

## Estimation and comparison of conventional and genomic breeding values in Holstein cattle of Antioquia, Colombia

### Estimación y comparación de valores de cría convencionales y genómicos en ganado Holstein de Antioquia, Colombia

Juan Zambrano A,<sup>1\*</sup> M.Sc, Juan Rincón F,<sup>2</sup> M.Sc, Albeiro López H,<sup>2</sup> Dr.Sci, Julián Echeverri Z,<sup>2</sup> Ph.D.

<sup>1</sup>Universidad Nacional de Colombia sede Medellín, Facultad de Ciencias, Posgrado en Biotecnología, Grupo de Investigación BIOGEM, A.A. Nro. 3840, Medellín, Colombia. <sup>2</sup>Universidad Nacional de Colombia sede Medellín, Facultad de Ciencias Agropecuarias, Departamento de producción Animal, Grupo de Investigación BIOGEM, A.A. Nro. 1779, Medellín, Colombia.\*Correspondencia: jczambranoa@unal.edu.co

Received: July 2014; Accepted: February 2015.

#### ABSTRACT

**Objective.** To estimate and compare breeding values (EBV) using the conventional method (BLUP) and genomic breeding values (MEBV and GEBV) estimated through bayes C method for milk yield and milk quality traits in dairy cattle in Antioquia, Colombia. **Materials and methods.** Two methods were used to estimate breeding values: BLUP to estimate conventional breeding value (EBV) and bayes C to estimate genomic values (MEBV and GEBV). The traits evaluated were: milk yield (PL), protein percentage (PPRO), fat percentage (PGRA) and score somatic cell (SCS). The methods (BLUP and bayes C) were compared using Person correlation ( $r_p$ ), Spearman rank correlation ( $r_s$ ) and linear regression coefficient (b). **Results.** The Pearson and Spearman correlations among EBVs and genomic values (MEBV and GEBV) ( $r_{p_{MEBV;EBV}}$  and  $r_{s_{GEBV;EBV}}$ ) were greater than 0.93 and the linear regression coefficients of EBVs on genomic values (MEBV and GEBV) ( $b_{MEBV;EBV}$  and  $b_{GEBV;EBV}$ ) ranged between 0.954 and 1.051 in all traits evaluated. **Conclusions.** The predictions of genomic values (MEBV and GEBV), using bayes C method were consistent with the predictions of the EBVs estimate through the conventional method (BLUP) in conditions of high Colombian tropic, allowing to obtain high associations between the breeding values.

**Key words:** BLUP, genotypes, genomic selection, single nucleotide polymorphism (Source: CAB, NAL).

#### RESUMEN

**Objetivo.** Estimar y comparar valores genéticos (EBV) usando el método convencional (BLUP) y valores genómicos (MEBV y GEBV) mediante el método bayes C en características de producción y calidad de la leche en ganado Holstein de Antioquia, Colombia. **Materiales y métodos.** Fueron empleados dos métodos para estimar valores genéticos: BLUP para estimar valores genéticos (EBV) y Bayes C para estimar valores genómicos (MEBV y GEBV). Las características evaluadas fueron: producción de leche (PL), porcentaje de proteína (PPRO), porcentaje de grasa (PGRA) y puntaje de células somáticas (SCS). Los métodos BLUP y bayes C fueron comparadas usando correlación de Pearson ( $r_p$ ), correlación por rangos de Spearman ( $r_s$ ) y regresión lineal (b). **Resultados.** Las correlaciones de Pearson y Spearman entre los EBVs y los valores genómicos (MEBV y GEBV) ( $r_{p_{MEBV;EBV}}$  y  $r_{s_{GEBV;EBV}}$ ) fueron mayores de 0.93 y los coeficientes de regresión entre los EBVs y los valores genómicos (MEBV y GEBV) ( $b_{MEBV;EBV}$  y

$b_{GEBV,EBV}$ ) oscilaron entre 0.954 y 1.051 en todas las características evaluadas. **Conclusiones.** La predicción de valores genómicos (MEBV y GEBV) usando el método Bayes C fue consistente con los EBVs estimados mediante el método BLUP en condiciones del trópico alto colombiano, permitiendo obtener altas asociaciones entre los valores genéticos.

**Palabras clave:** BLUP, genotipos, polimorfismo de nucleótido simple, selección genómica (*Fuente:* CAB, NAL).

## INTRODUCTION

The genetic improvement of plants and domestic animals includes among its stages, the prediction of estimated breeding values (EBV) with the objective of improving the genetic potential of next generations. Traditionally, EBVs are obtained by using the best linear unbiased prediction (BLUP) (1). This method is based on assumptions of the infinitesimal model, which assumes that the quantitative traits are determined by an infinite number of loci that act in an additive form, with no linkage among them. Each one with a very small effect on phenotypical expression of the traits (2,3). Although the BLUP method has been widely used for genetic progress of domestic species, its main limitation is the strict dependency on productive records to obtain high reliability in genetic evaluations (4).

Advances in genetics, molecular biology and other scientific areas have demonstrated that the quantitative traits are affected by an infinite number of loci, some with great effect, others with small effect and others with no effect on the traits (infinitesimal model) (3,5,6). With this approach, the termed marker-assisted selection (MAS) was initially incorporated in commercial applications, but with little success since it was directed to the use of individual genes or a small amount of QTL (quantitative trait loci) linked to markers that allow to capture only a small part of the genetic variance (7).

Thanks to the development of DNA sequencing technologies, a large number of genetic markers have been identified, being the single nucleotide polymorphism (SNP) among the highest. Currently, several types of microarrays allow for the identification of thousands of SNPs and are used in genotyping of some domestic species that are commercially available and at low cost (8). These microarray that include multiple SNP panels and cover all the genome in a uniform distribution, have been of great use in animal genotyping (9). Through the genomic selection methodology (SG) devised by Meuwissen et al (6), genomic information from genotyped animals with multiple SNPs can be used efficiently to estimate genomic values (GEBVs), which are SG's main point, given that selection decisions are based on these values. The genomic value

## INTRODUCCIÓN

El mejoramiento genético de plantas y animales domésticos incluye dentro de sus etapas, la predicción de valores estimados de cría (EBV), con el objetivo de mejorar el potencial genético de las siguientes generaciones. Tradicionalmente los EBVs, son obtenidos usando el mejor estimador lineal insesgado (BLUP) (1). Este método está basado en los supuestos del modelo infinitesimal, el cual asume que las características cuantitativas son determinadas por un número infinito de loci que actúan en forma aditiva, sin ligamiento entre ellos, cada uno con un efecto muy pequeño sobre la expresión fenotípica de la característica (2, 3). Si bien, el método BLUP ha sido ampliamente usado en el mejoramiento genético de especies domésticas, su principal limitación es la estricta dependencia de registros productivos para obtener altas confiabilidades en las evaluaciones genéticas (4).

Con los avances de la genética, la biología molecular y otras áreas de la ciencia, se ha demostrado que las características cuantitativas son afectadas por un número finito de loci, algunos con gran efecto, otros con efecto pequeño y otros sin efecto sobre la característica (modelo finitesimal) (3,5,6). Desde este enfoque, la selección asistida por marcadores moleculares (MAS) fue inicialmente incorporada en aplicaciones comerciales, pero con poco éxito, ya que fue orientada al uso de genes individuales o unos pocos QTL (loci de caracteres cuantitativos) ligados a marcadores, que permiten capturar sólo una pequeña proporción de la varianza genética (7).

Gracias al desarrollo de las tecnologías de secuenciación de ADN, se ha logrado la identificación de un gran número de marcadores genéticos y entre los más abundantes, los polimorfismos de nucleótido simple (SNP). En la actualidad, varios tipos de microarreglos que permiten la identificación de miles de SNPs y que son utilizados en la genotipificación de algunas especies domésticas, están disponibles comercialmente y a bajo costo (8). Estos microarreglos que incluyen paneles de múltiples SNPs y que cubren uniformemente todo el genoma, han sido de gran utilidad en la genotipificación de animales (9). Mediante

(GEBV), is the result of the sum of the effects of allelic substitution of all markers (SNP) associated with a phenotypic trait. These effects are estimated in selected populations that carry phenotypic and genealogical information and are later used in the estimation of GEBVs in young animals (8). Considering these arguments, several genetic value estimation methods that include molecular markers have been proposed, such as bayes A, bayes B, bayes C, bayes Cpi, bayes lasso, bayes R, GBLUP. The main difference between these methods is in the a priori information included in the genomic model with respect to the distribution of the marker's effects (6,10-12).

A big advantage in incorporating genomic information in conventional genetic evaluations is the increase in the reliability of genetic values in young animals and the decrease of the generation interval (13). This has led to many genetic improvement programs developed in several countries worldwide to include genomic information in their genetic evaluations; this strategy has worked and has allowed doubling the genetic gain in many characteristics of economic importance (8,14).

In Colombia, few genetic evaluations have been developed in dairy cattle. But due to the low number of animals evaluated and the limited amount of phenotypic observations, the reliability of EBVs has been low. Genomic evaluations have not yet been incorporated into dairy cattle, although some previous studies have been conducted (15). The objective of this research was to estimate and compare genetic values (EBV) using the conventional method (BLUP) and genomic breeding value (MEBV and GEBV) estimated through bayes C method for milk yield and milk quality traits in dairy cattle in Antioquia, Colombia.

## MATERIALS AND METHODS

**Population.** This research was conducted based on information from 124 dairy herds of Holstein breed located in 18 municipalities of the department of Antioquia (Colombia), which covers very wet low mountain (bmlh-mb) areas, with a temperature that ranges between 12 and 18°C and a yearly average of rain between 2000 and 4000 mm. These dairy herds are normally located at an altitude of between 1800 to 2800 msnm above sea level. The specific conditions of handling, feeding and sanitary, as well as their topography and geographic location varied in all dairy herds. The Ethics Research Committee of the National University of Colombia in Medellín supported this research, because it complied with

la metodología de selección genómica (SG), ideada por Meuwissen et al (6), puede utilizarse eficientemente la información genómica de animales genotipificados con múltiples SNPs, para estimar valores genómicos (GEBVs), que son el punto central de la SG, dado que las decisiones de selección son basadas en estos valores. El valor genómico (GEBV), es el resultado de la suma de los efectos de sustitución alélica de todos los marcadores (SNP) asociados a una característica fenotípica. Estos efectos, son estimados en poblaciones de referencia que portan información fenotípica y genealógica y que luego son utilizados en la estimación de GEBVs en animales jóvenes (8). Considerando estos argumentos, varios métodos de estimación de valores genéticos que incluyen marcadores moleculares han sido propuestos, como bayes A, bayes B, bayes C, bayes Cpi, bayes lasso, bayes R, GBLUP, entre otros y la principal diferencia entre estos métodos se da en la información a priori incluida en el modelo genómico, con respecto a la distribución de los efectos de los marcadores (6,10-12).

La gran ventaja de incorporar la información genómica en evaluaciones genéticas convencionales es el incremento de la confiabilidad de los valores genéticos en animales jóvenes y la disminución del intervalo generacional (13). Esto, ha llevado a que muchos programas de mejoramiento genético desarrollados en varios países alrededor del mundo, incluyan información genómica en sus evaluaciones genéticas; estrategia que ha funcionado y que ha permitido doblar la tasa de ganancia genética en muchas características de importancia económica (8,14).

En Colombia han sido desarrolladas escasas evaluaciones genéticas, en ganado de leche, pero debido al bajo número de animales evaluados y a la limitada cantidad de observaciones fenotípicas, las confiabilidades de los EBVs han sido bajas. En cuanto a evaluaciones genómicas, aún no han sido incorporadas en ganado de leche, aunque han sido realizados algunos estudios previos (15). El objetivo de esta investigación fue estimar y comparar valores genéticos (EBV) usando el método convencional (BLUP) y valores genómicos (MEBV y GEBV) mediante el método bayes C en características de producción y calidad de la leche en ganado Holstein de Antioquia, Colombia.

## MATERIALES Y MÉTODOS

**Población.** Esta investigación fue realizada con base en la información de 124 hatos lecheros de raza Holstein ubicados en 18 municipios del departamento de Antioquia (Colombia) que cubren zonas de bosque muy húmedo montano bajo

the standards for this type of research (Approval number: CEMED-015 May, 2012).

The number of records used for the analysis varied according to the dependent variables analyzed: 8772, 6624, 6595 and 6426 for PL, PGRA, PPRO y and somatic cell count (RCS) respectively. RCS was transformed to a somatic cell score (SCS) using the following equation:  $SCS = [\log_2(RCS/100000)] + 3$ , in order to improve the normality of the data, as described by Ali and Shook (16). The number of animals in the pedigree for the conventional genetic evaluation was 9090 individuals and for the genomic evaluation the number was 260, which corresponds to the genotyped animals and their relatives. All the information was administered and analyzed in Control 1 Software, version 1.0 (17).

According to the initial descriptive analysis, PL values greater than 12000 and less than 1500 liters/lactation were discarded. Levels lower than 2.00% and greater than 5.00% for PPRO and less than 1.60% and greater than 6.00% for PGRA were also discarded since they are considered possible sampling errors or errors in the lab's procedures.

**Animal genotyping with high-density beadchips.** 144 animals were genotyped with Illumina (Illumina Inc., San Diego, CA) beadchips for multiple SNPs: 31 with the Bovine LD beadchip, 65 with the Bovine SNP50v2 beadchip and 48 with the Bovine HD beadchip, which provide information of 6.909, 54.609 and 777.972 SNPs respectively. During the handling of the genotypes, the alleles with the lowest frequency were removed when these were lower than 0.02. Genotypes with mendelian errors were also declared as lost data. Finally, after editing the information, 6716 SNPs common to all three types of beadchips (LD, SNP50 and HD) were obtained for all the genotyped animals. The edition of genotypes was conducted with the programs SAS v9.1 (18) and Plink v1.07 (19).

#### Statistical analysis.

**EBVs Estimation by BLUP.** Through preliminary analysis, several statistical models were proven that included significant fixed effects for each trait in order to determine the models that were best adjusted to the estimation of the genetic parameters. variance components and heritability were estimated using an independent univariate animal model for each characteristic (20). The matrix notation of the model was described as follows:

$$y = Xb + Za + e$$

Where:  $y$  = observation vector,  $b$  = solution vector for fixed effects of the herd, number

( $b_{mh-mb}$ ), con una temperatura que oscila entre 12 y 18°C y un promedio anual de lluvias entre 2000 y 4000 mm. Normalmente se extienden en una faja altimétrica de 1800 a 2800 msnm. Las condiciones específicas de manejo, alimentación y sanidad fueron variables en todos los hatos, así como su topografía y ubicación geográfica. Esta investigación fue avalada por el Comité de Ética en Investigación de la Universidad Nacional de Colombia sede Medellín, considerando que cumplió con los estándares para este tipo de investigación (Número de la carta de aprobación: CEMED-015 Mayo, 2012).

El número de registros utilizados para los análisis fue variable, de acuerdo con la característica dependiente analizada: 8772, 6624, 6595 y 6426 para PL, PGRA, PPRO y recuento de células somáticas (RCS) respectivamente. El RCS fue transformado a puntaje de células somáticas (SCS) mediante la siguiente ecuación:  $SCS = [\log_2(RCS/100000)] + 3$ , con el fin de mejorar la normalidad de los datos, como lo describe Ali y Shook (16). El número de animales en el pedigrí para la evaluación genética convencional fue de 9090 individuos y para evaluación genómica fue de 260, que corresponde a los animales genotificados y sus parientes. La totalidad de la información fue administrada y analizada en el Software Control 1, versión 1.0 (17).

De acuerdo con el análisis descriptivo inicial, valores de PL superiores a 12000 e inferiores a 1500 litros/lactancia fueron descartados, niveles inferiores a 2.00% y superiores a 5.00% para PPRO y menores a 1.60% y mayores a 6.00% para PGRA igualmente fueron descartados, por considerarse posibles errores de muestreo o errores de procedimientos de laboratorio.

**Genotificación de animales con chips de alta densidad.** Fueron genotificados 144 animales con chips de Illumina (Illumina Inc., San Diego, CA) para múltiples SNPs: 31 con el chip Bovine LD, 65 con el chip Bovine SNP50v2 y 48 con el chip Bovine HD, los cuales proveen información de 6.909, 54.609 y 777.972 SNPs respectivamente. En la edición de los genotipos, se removieron los alelos de menor frecuencia cuando esta era inferior a 0.02, también fueron declarados como datos perdidos los genotipos con errores mendelianos. Finalmente, después de editar la información, se obtuvieron 6716 SNPs comunes a los tres chips (LD, SNP50 y HD) para todos los animales genotificados. La edición de los genotipos fue realizada con los programas SAS v9.1 (18) y Plink v1.07 (19).

#### Análisis estadístico.

**Estimación de EBVs usando BLUP.** Mediante

of calving (cows between 1 and 8 calving), covariates: lactation length for PL and milk yield (PL) for PPRO and PGRA traits,  $a$  = solution vector for random additive genetic effect,  $X$  and  $Z$  = incidence matrix that relate the records with fixed and random effects and  $e$  = residual.

It is assumed that the random and residual additive genetic effects are distributed independently with zero medium and variances:

$$V \begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

Considering that  $A$  = relationship matrix and:  $I\sigma_e^2 = R$ , you have to:

$$\text{var}(y) = ZAZ'\sigma_a^2 + R$$

The mixed model equations for the best linear unbiased estimator (BLUE) of the estimated functions for  $\mathbf{b}$  for the best linear unbiased predictor (BLUP) of  $\mathbf{a}$  are describe as:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Where:  $\alpha$  is the relationship between the residual variance and the additive genetic variance:  $\alpha = \sigma_e^2 / \sigma_a^2$ .

The model for each characteristic was resolved by using the MTDFREML program described by Boldman et al (21), which determined the solutions for each fixed effect ( $b$ ) and the additive genetic random effect ( $a$ ) from the mixed model equations (MME), described by Henderson (1), by using a multiple trait derivative-free restricted maximum likelihood.

The traits were estimated directly from the MTDFREML program and were considered as the proportion of the additive genetic variance over the phenotype variance.

**Estimation of marker effects and genomic values.** The estimation of the effects of the SNPs and the genomic breeding values (MEBV and GEBV) was conducted using the Bayes C method (22). The general statistical model is describe as:

$$y = 1_n \mu + \sum_{j=1}^p X_{ij} g_j \delta_j + Zu + e$$

Where:  $y$  is the phenotype value vector corrected for fixed effects (corresponding to EBVs obtained through the BLUP method) for PL, SCS, PPRO, and PGRA traits,  $\mu$  is the general mean,  $1_n$  is the

análisis preliminares se probaron varios modelos estadísticos que incluyeron efectos fijos significativos para cada característica, con el fin de determinar los modelos que mejor se ajustaban en la estimación de los parámetros genéticos. Los componentes de varianza y la heredabilidad fueron estimados usando un modelo animal univariado independiente para cada una de las características (20). La notación matricial del modelo fue descrita de la siguiente manera:

$$y = Xb + Za + e$$

Donde:  $y$  = vector de observaciones,  $b$  = vector de solución para los efectos fijos hato, número de parto (vacas entre 1 y 8 partos), covariables duración de la lactancia para PL, y producción de leche (PL) para las características PPRO y PGRA,  $a$  = vector de soluciones para el efecto aleatorio genético aditivo,  $X$  y  $Z$  = matrices de incidencia que relacionan los registros con los efectos fijos y aleatorios y  $e$  = residual.

Se asume que los efectos aleatorios genético aditivo y residual son distribuidos independientemente con media cero y varianzas:

$$V \begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

Considerando que  $A$  = matriz de parentesco y  $I\sigma_e^2 = R$ , se tiene que:

$$\text{var}(y) = ZAZ'\sigma_a^2 + R$$

Las ecuaciones del modelo mixto para el estimador lineal insesgado (BLUE) de las funciones estimables de  $\mathbf{b}$  y para el mejor predictor lineal insesgado (BLUP) de  $\mathbf{a}$  se describen como:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Donde:  $\alpha$  es la relación entre la varianza del error y la varianza genética aditiva:  $\alpha = \sigma_e^2 / \sigma_a^2$ .

El modelo para cada característica fue resuelto usando el programa MTDFREML descrito por Boldman et al (21) y que determina las soluciones para cada efecto fijo ( $b$ ) y efecto aleatorio genético aditivo ( $a$ ) de las ecuaciones del modelo mixto (MME), descrito por Henderson (1), empleando el método de máxima verosimilitud restringida libre de derivadas.

La heredabilidad se estimó directamente del programa MTDFREML y se consideró como la

vector of ones of  $n$  length,  $X_{ij}$  is the column vector that represents the covariates of the marker (SNP) in locus  $j$  and animal  $i$ . In each  $j$ th marker, there are three possible combinations of two alleles (A or B), the homozygote of an allele (AA), the heterozygote (AB) and the homozygote of the other allele (BB). These combinations of alleles (genotypes) are represented quantitatively as 0, 1 and 2 respectively (that is;  $X_{ij}=0, 1$  o  $2$ );  $g_j$  is the random effect of allelic substitution for the marker  $j$ , which is conditional on  $\sigma_g^2$  and is normally distributed  $N(0, \sigma_g^2)$  when  $\delta_j=1$ , but  $g_j=0$  when  $\delta_j=0$ ;  $\delta_j$  is a random variable 0/1 that indicates absence (with probability  $\Pi$ ) or presence (with probability  $1-\Pi$ ) of locus  $j$  in the model;  $u$  is the polygenic random effect vector of length  $n$  ( $Z$  is the related design matrix) and can be considered as the adjustment of the genes not considered due to the effects of  $j$ th SNPs in  $g$ ;  $u$  is assumed to be normally distributed,  $u \sim N(0, A\sigma_u^2)$ , where  $A$  is the relationship matrix derived from the pedigree of the genotyped animals;  $e$  is the residual and also is assumed to be normally distributed  $e \sim N(0, I\sigma_e^2)$ , where  $I$  is an  $n \times n$  identity matrix.

With this approach, the bayes C method assumes a combination of two distributions for the effects of the marker SNPs. The first distribution, (with  $\pi$  probability) established that many markers do not have effect nor variance; and the second distribution, (with  $1-\pi$  probability) there are markers with effect and variance, which common for all SNPs (23).

Therefore, the effects of SNPs ( $g_j$ ) follow an apriori normal distribution with zero mean and variance  $\sigma_g^2$ , while the variances of the effects of the markers ( $\sigma_g^2$ ) follow an apriori inverse scaled chi squared-distribution with parameters  $V_g$  (degrees of freedom) and  $S_g^2$  (scale parameters). On the other hand,  $\pi$  is treated as a uniform apriori distribution with zero mean and variance 1 (23), as follows:

$$g_j | \pi, \sigma_g^2 \begin{cases} \sim N(0, \sigma_g^2) \rightarrow \text{con probabilidad } 1-\pi \\ = 0 \rightarrow \text{con probabilidad } \pi \end{cases}$$

$$\sigma_g^2 | V_g S_g^2 \sim V_g S_g^2 \chi^2$$

$$\pi \sim \text{uniforme}(0,1)$$

The genomic value of animals whose genotype is known was determined as:

$$y = \sum_{j=1}^p X_{ij} \hat{g}_j \delta_j + \hat{\mu}$$

Where:  $y$  is the genomic estimated breeding value (GEBV),  $\sum_{j=1}^p X_{ij} \hat{g}_j \delta_j$  corresponds to the estimated molecular breeding value (MEBV)

proporción de la varianza genética aditiva sobre la varianza fenotípica.

**Estimación de los efectos de los marcadores y valores genómicos.** La estimación de los efectos de los SNPs y de los valores de cría genómicos (MEBV y GEBV) fue realizada empleando el método Bayes C (22). El modelo estadístico general se describe como:

$$y = \mathbf{1}_n \mu + \sum_{j=1}^p X_{ij} g_j \delta_j + Zu + e$$

Donde:  $y$  es el vector de valores fenotípicos corregidos por efectos fijos (que corresponde a los EBVs obtenidos mediante el método BLUP) para las características PL, SCS, PPRO y PGRA,  $\mu$  es la media general,  $\mathbf{1}_n$  es el vector de unos de longitud  $n$ ,  $X_{ij}$  es el vector columna que representa la covariable del marcador (SNP) en el locus  $j$  en el animal  $i$ . En cada  $j$ th marcador son tres posibles combinaciones de dos alelos (A o B), el homocigótico de un alelo (AA), el heterocigótico (AB) y el homocigótico del otro alelo (BB). Estas combinaciones de alelos (genotipos) son representados cuantitativamente como 0, 1 y 2 respectivamente (es decir;  $X_{ij}=0, 1$  o  $2$ ),  $g_j$  es el efecto aleatorio de sustitución alélica para el marcador  $j$ , el cual es condicional sobre  $\sigma_g^2$  y se asume normalmente distribuido

$N(0, \sigma_g^2)$  cuando  $\delta_j=1$ , pero  $g_j=0$  cuando  $\delta_j=0$ ,  $\delta_j$  es una variable aleatoria 0/1 que indica ausencia (con probabilidad  $\Pi$ ) o presencia (con probabilidad  $1-\Pi$ ) del locus  $j$  en el modelo,  $u$  es el vector de efectos aleatorios poligénicos de longitud  $n$  ( $Z$  es la matriz de diseño asociada) y puede ser considerado como el ajuste de los genes no tenidos en cuenta por los efectos de los  $j$ th SNPs en  $g$ ,  $u$  es asumido estar distribuido normalmente,  $u \sim N(0, A\sigma_u^2)$ , donde  $A$  es la matriz de parentesco derivada del pedigrí de los animales genotipificados,  $e$  es el residual y también se asume estar distribuido normalmente,  $e \sim N(0, I\sigma_e^2)$ , donde  $I$  es una matriz identidad  $n \times n$ .

De esta manera, el método bayes C asume una mezcla de dos distribuciones para los efectos de los SNPs. En la primera, (con probabilidad  $\pi$ ) establece que muchos marcadores no tienen efecto ni varianza y para la segunda, (con probabilidad  $1-\pi$ ) existen marcadores con efecto y varianza, la cual es común para todos los SNPs (23).

Así, los efectos de los SNPs ( $g_j$ ) siguen una distribución apriori normal con media cero y varianza  $\sigma_g^2$ , mientras que las varianzas de los efectos de los marcadores ( $\sigma_g^2$ ) siguen una distribución apriori chi cuadrada invertida escalada con parámetros  $V_g$  (grados de libertad)

obtained as the sum of all estimated effects ( $\hat{g}_j$ ) of the SNPs;  $\hat{u}$  is the polygenic effect.

The program used for the estimation of the effects of SNPs and the genomic values (MEBV and GEBV) was GS3 (24).

**Methods for comparing genetic values.**

Genomic values (MEBV and GEBV) estimated with the bayes C method were compared to conventional genetic values (EBV), estimated with the BLUP method, using three methods: Spearman rank correlation, Pearson correlation and simple linear regression:

a) Spearman's rank correlation coefficient was used to determine the degree of similarity between the classifications or rankings of the animals by their genomic values (MEBV or GEBV) and the classification of the same animals by the conventional EBV. The equation used was:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Where:

D = difference between the corresponding genetic values of (x-y) order,  
 x = genomic value (GEBV or MEBV),  
 y = conventional EBV,  
 N = number of pairs (x,y) and  
 $r_s$  = Spearman correlation coefficient.  
 A value of correlation 1 indicates that the classification or ranking of the animals by their genomic values (MEBV or GEBV) is similar than in the conventional EBV.

b) The Pearson correlation coefficient was used to determine the relationship between genomic values (MEBV or GEBV) and the conventional EBV. The equation used was:

$$r_p = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Where,  $r_p$  = Pearson correlation coefficient;  
 x=genomic value (GEBV or MEBV) for i trait;  
 y=conventional EBV for i trait. A correlation value of 1 indicates that the methods for estimation of genetic values have the same precision and that the estimation errors are very similar.

c) The simple linear regression coefficient was used to determine the change in magnitude of the genomic values (MEBV and GEBV) with respect to conventional EBVs and allow evaluate the bias in the estimation of the genomic selection model with respect to the traditional model, according to Mäntysaari et al (25). Linear regression was defined as:

$y | S_g^2$  (parámetro de escala). Por otra parte,  $n$  es tratada como una distribución apriori uniforme con media cero y varianza 1 (23), de la siguiente manera:

$$g_j | \pi, \sigma_g^2 \begin{cases} \sim N(0, \sigma_g^2) \rightarrow \text{con probabilidad } 1 - \pi \\ = 0 \rightarrow \text{con probabilidad } \pi \end{cases}$$

$$\sigma_g^2 | \nu_g S_g^2 \sim \nu_g S_g^2 x^{-2}$$

$$\pi \sim \text{uniforme}(0,1)$$

El valor genómico de los animales cuyo genotipo es conocido, se determinó como:

$$y = \sum_{j=1}^p X_{ij} \hat{g}_j \delta_j + \hat{u}$$

Donde:  $y$  es el valor genómico (GEBV),  $\sum_{j=1}^p X_{ij} \hat{g}_j \delta_j$  corresponde al valor estimado de cría molecular (MEBV) obtenido como la suma de todos los efectos estimados ( $\hat{g}_j$ ) de los SNPs,  $\hat{u}$  es el efecto poligénico.

El programa utilizado para la estimación de los efectos de los SNPs y de los valores genómicos (MEBV y GEBV) fue GS3 (24).

**Metodologías de comparación de valores genéticos.**

Los valores genómicos (MEBV y GEBV) estimados mediante el método bayes C fueron comparados con los valores genéticos convencionales (EBV), estimados mediante el método BLUP, usando tres métodos: correlación por rangos de Spearman, correlación de Pearson y regresión lineal simple:

a) El coeficiente de correlación por rangos de Spearman fue usado para determinar el grado de similitud entre las clasificaciones o ranking de los animales por sus valores genómicos (MEBV o GEBV) y clasificación de los mismos animales por el EBV convencional. La ecuación empleada fue:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Dónde:

D = diferencia entre los correspondientes valores genéticos de orden (x-y),  
 x = valor genómico (GEBV o MEBV),  
 y = EBV convencional,  
 N = número de parejas (x,y) y  
 $r_s$  = coeficiente de correlación de Spearman.  
 Un valor de correlación de 1 indica que la clasificación o ranking de los animales por sus valores genómicos (MEBV y GEBV) es similar en los EBVs convencionales.

b) El coeficiente de correlación de Pearson fue usado para determinar la asociación entre los valores genómicos (MEBV o GEBV) y el EBV convencional. La ecuación empleada fue:

$$y_i = b + b_1x_i + e$$

Where  $y_i$  = dependent variable that corresponds to conventional EBV for  $i$  trait;  $x_i$ =genomic value (GEBV or MEBV) for  $i$  characteristic;  $b_1$ =EBV regression coefficient on the genomic value (GEBV or MEBV);  $b_0$  = intercept and  $e$ =residual.

## RESULTS

**Descriptive analysis.** The averages for the milk yield and quality traits presented normal values for the Holstein breed in the Colombian high tropic, with values of 5400 liters/lactation, 3.10%, 3.97% and 4.72 for PL, PPRO, PGRA and SCS respectively. The most variable trait according to the coefficient of variation was PL with 39.9%, while the trait with the lowest variance was PPRO with 9.9% (Table 1). The effects of the SNPs for the same trait were also determined. For PL, the effects varied between -43.3 and 34.5 liters/lactation; for PPRO between -0.019 and 0.022%; for PGRA between -0.022 and 0.025% and for SCS between -0.055 and

**Table 1.** Descriptive analysis for milk yield (PL), protein percentage (PPRO), fat percentage (PGRA) and score somatic cells (SCS) in Holstein cattle in Antioquia, Colombia.

Traits	N	Mean±DE	CV(%)
PL(liters/lactation)	670	5400±2126	39.4
PPRO(%)	457	3.10±0.31	9.9
PGRA(%)	455	3.97±0.55	13.9
SCS	452	4.62±1.37	29.7

N: Number of records, DE: Standard deviation, CV: Coefficient of variation.

**Table 2.** Descriptive statistical analysis of the effects of SNP's for milk yield (PL), protein percentage (PPRO), fat percentage (PGRA) and score somatic cells (SCS) in Holstein cattle in Antioquia, Colombia.

Traits	N	Mean±DE	EE	Minimum	Maximum
PL (liters/lactation)	6716	-0.033 ±5.49	8.23	-43.3	34.5
PPRO(%)	6716	0.00003 ±0.002	0.003	-0.019	0.022
PGRA(%)	6716	0.00005 ±0.003	0.003	-0.022	0.025
SCS	6716	0.00008 ±0.007	0.007	-0.055	0.054

N: Number of SNPs, DE: Standard deviation, EE: Average standard error for SNP effects.

0.054 (Table 2).

**Genetic parameters, conventional EBVs and genomics (MEBV and GEBV).** The heritability estimates for the bayes C method for PGRA and SCS were within the normal range and were

$$r_p = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Donde,  $r_p$  = coeficiente de correlación de Pearson,  $x$  = valor genómico (GEBV o MEBV) para la característica  $i$ ,  $y$  = EBV convencional para la característica  $i$ . Un valor de correlación de 1 indica que los métodos de estimación de valores genéticos tienen la misma precisión y que los errores de estimación son muy similares.

c) El coeficiente de regresión lineal simple fue usado para determinar el cambio en magnitud de los valores genómicos (MEBV y GEBV) con respecto a los EBVs convencionales y permite evaluar e sesgo en la estimación del modelo de selección genómica con respecto al modelo tradicional, según Mäntysaari et al (25). La regresión lineal fue definida como:

$$y_i = b + b_1x_i + e$$

Donde  $y_i$  = variable dependiente que corresponde EBV convencional para la característica  $i$ ,  $x_i$ =valor genómico (GEBV o MEBV) para la característica  $i$ ,  $b_1$ =coeficiente de regresión del EBV convencional sobre el valor genómico (GEBV o MEBV),  $b_0$ =intercepto y  $e$ =residual.

## RESULTADOS

**Análisis descriptivo.** Las medias para las características productivas y de calidad composicional de la leche presentaron valores normales para la raza Holstein en el trópico alto colombiano, con valores de 5400 litros/lactancia, 3.10%, 3.97% y 4.72 para PL, PPRO, PGRA y SCS respectivamente. La característica más variable según el coeficiente de variación fue PL con 39.9%, mientras que la característica que presentó menor variabilidad fue PPRO con 9.9% (Tabla 1). Los efectos de los SNPs para las mismas características también fueron determinados. Para PL los efectos oscilaron entre -43.3 y 34.5 litros/lactancia, para PPRO entre -0.019 y 0.022%, para PGRA entre -0.022 y 0.025% y para SCS entre -0.055 y 0.054 (Tabla 2).

**Parámetros genéticos, EBVs convencionales y genómicos (MEBV y GEBV).** Las heredabilidades estimadas por el método bayes C para PGRA y SCS estuvieron dentro del rango normal y fueron cercanas a las estimadas por el método BLUP. Sin embargo, la heredabilidad para PPRO fue alta, comparada con las heredabilidades obtenidas por el método BLUP. Las medias de los valores genéticos (EBV) obtenidos por BLUP para PL, PPRO, PGRA y SCS fueron: 138 litros/lactancia, 0.002%, -0.011% y -0.019



close to the heritabilities estimates by the BLUP method. But, the heritability for PPRO was high compared to the heritability obtained with the BLUP method. The averages of the genetic values (EBV) obtained by BLUP for PL, PPRO, PGRA and SCS were: 138 liters/lactation, 0.002%, -0.011% and -0.019 respectively. For GEBVs, the averages were: -38 liters/lactation, -0.008%, -0.012% and -0.023 and for MEBVs, which were calculated as the sum of the effects of the SNPS, the averages

**Table 3.** Heritabilities, conventional genetic value (EBVs) mean and genomic values (MEBV and GEBV) mean for milk yield (PL), protein percentage (PPRO), fat percentage (PGRA) and score somatic cells (SCS) in Holstein cattle in Antioquia, Colombia.

Trait	h <sup>2</sup> _BLUP	h <sup>2</sup> _bayesC	EBV±DE	MEBV±DE	GEBV±DE
PL	0.30 ±0.06	0.13	138 ±558	-51 ±522	-38 ±549
PPRO	0.33 ±0.02	0.64	0.002 ±0.089	-0.01 ±0.087	-0.008 ±0.087
PGRA	0.41 ±0.02	0.31	-0.011 ±0.153	-0.015 ±0.146	-0.012 ±0.154
SCS	0.27 ±0.02	0.25	-0.019 ±0.311	-0.025 ±0.319	-0.023 ±0.314

DE: standard deviation, h<sup>2</sup>\_BLUP: traits obtained using BLUP, h<sup>2</sup>\_bayesC: traits obtained by the bayes C method.

were: 51 liters/lactation, -0.01%, -0.015% and -0.025 respectively (Table 3).

**Comparison of genetic values.** Pearson correlations between genomic values (GEBV and MEBV) and the conventional EBV ( $r_{pMEBV;EBV}$ ,  $r_{pGEBV;EBV}$ ) for PL, PPRO, PGRA y SCS were greater than 0.94 and Spearman correlations ( $r_{sMEBV;EBV}$ ,  $r_{sGEBV;EBV}$ ) obtained for the same trait were greater than 0.93 and in most cases close to 1, indicating high relationship between genetic values. In all cases, the correlations obtained were highly significant ( $p \leq 0.01$ ), (Table 4).

Regression coefficients ( $b_{MEBV;EBV}$  y  $b_{GEBV;EBV}$ ) obtained for PL, PPRO, PGRA y SCS varied between 0.954 and

**Table 4.** Pearson correlation coefficients and Spearman correlation coefficients between genomic values (GEBV and MEBV) and conventional EBV's.

Trait	Pearson Correlation Coefficient		Spearman Correlation Coefficient	
	$r_{pMEBV;EBV} \pm EE$	$r_{pGEBV;EBV} \pm EE$	$r_{sMEBV;EBV} \pm EE$	$r_{sGEBV;EBV} \pm EE$
PL	0.984 ±0.0026	0.993 ±0.0012	0.981 ±0.0006	0.991 ±0.0005
PPRO	0.947 ±0.0086	0.989 ±0.0019	0.932 ±0.020	0.983 ±0.0040
PGRA	0.967 ±0.0055	0.995 ±0.0009	0.959 ±0.005	0.989 ±0.00008
SCS	0.979 ±0.0035	0.989 ±0.0018	0.972 ±0.029	0.986 ±0.0020

$r_{pMEBV;EBV}$ : Pearson correlation between MEBV and EBV,  $r_{pGEBV;EBV}$ : Pearson correlation between GEBV and EBV,  $r_{sMEBV;EBV}$ : Spearman correlation between MEBV and EBV,  $r_{sGEBV;EBV}$ : Spearman correlation between GEBV and EBV, EE: standard error.

respectivamente. Para los GEBVs, las medias fueron: -38 litros/lactancia, -0.008%, -0.012% y -0.023 y para los MEBV, que fueron calculados como la sumatoria de los efectos de los SNPs fueron: 51 litros/lactancia, -0.01%, -0.015% y -0.025 respectivamente (Tabla 3).

**Comparación de valores genéticos.** Las correlaciones de Pearson entre los valores genómicos (GEBV y MEBV) y el EBV convencional ( $r_{pMEBV;EBV}$ ,  $r_{pGEBV;EBV}$ ) para PL, PPRO, PGRA y SCS fueron superiores a 0.94 y las correlaciones de Spearman ( $r_{sMEBV;EBV}$ ,  $r_{sGEBV;EBV}$ ) obtenidas para las mismas características fueron superiores a 0.93 y en la gran mayoría muy cercanas a 1, indicando asociaciones altas entre valores genéticos. En todos los casos, las correlaciones obtenidas fueron altamente significativas ( $p \leq 0.01$ ), (Tabla 4).

Por su parte, los coeficientes de regresión ( $b_{MEBV;EBV}$  y  $b_{GEBV;EBV}$ ) obtenidos para PL, PPRO, PGRA y SCS oscilaron entre 0.954 y 1.051. Estos resultados indican que la estimación de valores genéticos en la población evaluada, usando los métodos BLUP y Bayes C no alteró significativamente la magnitud de los valores genómicos (GEBV y MEBV) con respecto al EBV convencional. De igual manera, los coeficientes de determinación ( $R^2$ ) para las características evaluadas (PL, PPRO, PGRA y SCS fueron cercanos a 1 (mayor de 0.89 para las regresiones lineales entre EBV y MEBV y mayor de 0.97 para las regresiones lineales entre EBV y GEBV), lo cual muestra una relación lineal alta entre los valores genéticos (los datos se ajustan a una línea recta), (Tabla 5).

## DISCUSIÓN

La selección genómica (SG), es una variante de la selección asistida por marcadores moleculares

**Table 5.** Regression coefficients between genomic values (GEBV and MEBV) and conventional EBVs for milk yield (PL), protein percentage (PPRO), fat percentage (PGRA) and score somatic cells (SCS) in Holstein cattle in Antioquia, Colombia.

Trait	Regression coefficient			
	$b_{MEBV;EBV} \pm EE$	$R^2$	$b_{GEBV;EBV} \pm EE$	$R^2$
PL	1.051 ±0.016	0.969	1.008 ±0.010	0.985
PPRO	0.960 ±0.027	0.898	1.002 ±0.012	0.978
PGRA	1.015 ±0.023	0.934	0.988 ±0.009	0.989
SCS	0.954 ±0.018	0.958	0.979 ±0.012	0.978

$b_{MEBV;EBV}$ : regression coefficient between MEBV and EBV,  $b_{GEBV;EBV}$ : regression coefficient between GEBV and EBV, EE: standard deviation,  $R^2$ : Coefficient of determination.

1.051. These results indicate that the estimation of genetic values in the population evaluated using BLUP and Bayes C methods, did not alter significantly the magnitude of the genomic values (GEBV and MEBV) with respect to the conventional EBV. Likewise, the determination coefficients ( $R^2$ ) for the trait evaluated (PL, PPRO, PGRA and SCS) were close to 1 (greater than 0.89 for linear regressions between EBV and MEBV and greater than 0.97 for linear regressions between EBV and GEBV). This shows a high linear relationship between genetic values (the data adjusts to a straight line), (Table 5).

## DISCUSSION

The genomic selection (SG), is a type of marker-assisted selection (MAS), but unlike other methodologies, includes molecular markers massively, which cover all the entire genome, in such a way that all QTL's are in linkage disequilibrium with at least one marker (5). SG allows a favorable selection of individuals based on the breeding genomic value (GEBV) and compared with conventional BLUP (1), provides breeding values more accurate in young animals; but the BLUP method provides breeding value more accurate in older animals that have progeny tests (26). Also, SG, unlike the BLUP method, can be more efficient in both low heritability as well as in high heritability traits (27,28).

This research determined that heritabilities for PL, PGRA, PPRO and SCS, through the BLUP and bayes C methods. Heritabilities for PL, PGRA y SCS were higher when the BLUP method was used with values between 0.30, 0.41 and 0.27, compared to the estimated using the bayes C method, with values of 0.13, 0.41 and 0.25 respectively. For PPRO, the heritability was higher using the bayes C method with a value of 0.64 and using the BLUP method the value was 0.33. VanRaden et al (13) estimated the heritabilities for milk production traits using genomic linear predictions and obtained values of 0.30 for PL, 0.50 for PPRO and PGRA and 0.12 for SCS. In both research, heritabilities are within normal ranges, except in the current research, traits for PPRO (estimated with bayes C), was two-times the heritability estimated with BLUP. According to Visscher et al (29), false associations can inflate the heritability when genomic methods are used.

In this research, conventional breeding values (EBV) were also estimated with the BLUP method and genomic values (MEBV and GEBV), obtained with the Bayes C method for milk productions and quality traits PL, PPRO, PGRA and SCS. The genomic values (MEBV and GEBV) were compared with conventional EBV using as

(MAS), pero a diferencia de otras metodologías MAS, incluye marcadores de forma masiva, que cubren todo el genoma, de tal manera que todos los QTL estén en desequilibrio de ligamiento con al menos un marcador (5). La SG permite la selección favorable de individuos basada en el valor de cría genómico (GEBV) y comparada con el BLUP convencional (1), provee valores de cría más exactos en animales jóvenes; pero el método BLUP provee valores de cría con mayor exactitud en animales adultos que van envejeciendo y que poseen pruebas de progenie (26). Además, la SG, a diferencia del método BLUP, puede ser más eficiente tanto en características de baja como de alta heredabilidad (27, 28).

En esta investigación se determinó la heredabilidad para PL, PGRA, PPRO y SCS, mediante los métodos BLUP y bayes C. Las heredabilidades para PL, PGRA y SCS fueron mayores cuando se usó el método BLUP con valores de 0.30, 0.41 y 0.27, comparadas con las estimadas usando bayes C, con valores de 0.13, 0.41 y 0.25 respectivamente. Para PPRO la heredabilidad fue mayor usando el método bayes C con un valor de 0.64 y usando BLUP fue 0.33. VanRaden et al (13), estimaron las heredabilidades para características productivas usando predicciones lineales genómicas y obtuvieron valores de 0.30 para PL, 0.50 para PPRO y PGRA y 0.12 para SCS. En las dos investigaciones las heredabilidades están dentro de los rangos normales, excepto que en la investigación actual, la heredabilidad para PPRO (estimada con bayes C), fue el doble de la estimada con BLUP. Según Visscher et al (29), falsas asociaciones pueden inflar la heredabilidad cuando se usan métodos genómicos.

En esta investigación, también fueron estimados valores de cría convencionales (EBV) mediante el método BLUP y valores genómicos (MEBV y GEBV), obtenidos mediante el método Bayes C, para las características productivas PL, PPRO, PGRA y SCS. Los valores genómicos (MEBV y GEBV), fueron comparados con el EBV convencional, usando como medida las correlaciones de Pearson ( $r_{p_{MEBV;EBV}}$  y  $r_{p_{GEBV;EBV}}$ ) y Spearman ( $r_{s_{MEBV;EBV}}$  y  $r_{s_{GEBV;EBV}}$ ) y el coeficiente de regresión lineal ( $b_{MEBV;EBV}$  y  $b_{GEBV;EBV}$ ). Las correlaciones de Pearson y Spearman entre el GEBV y el EBV fueron altas en todos los casos (mayores de 0.98) y los coeficientes de regresión lineal del EBV sobre el GEBV fueron cercanos a 1 (oscilaron entre 0.979 y 1.008) como se indica en las tablas 4 y 5. Estos coeficientes son elevados dado que se utiliza un modelo que combina un primer efecto aleatorio asociado con los marcadores y un segundo efecto denominado efecto poligénico que tiene en cuenta la matriz de parentesco de los animales genotipificados (BLUP

measure the Pearson correlations ( $r_{p_{MEBV;EBV}}$  and  $r_{p_{GEBV;EBV}}$ ) and Spearman correlations ( $r_{s_{MEBV;EBV}}$  and  $r_{s_{GEBV;EBV}}$ ) and the linear regression coefficient ( $b_{MEBV;EBV}$  and  $b_{GEBV;EBV}$ ). Pearson and Spearman correlations between the GEBV and the EBV were high in all cases (greater than 0.98) and linear regression coefficients of EBV on GEBV were close to 1 (ranged between 0.979 and 1.008) as indicated in tables 4 and 5. These coefficients are high since a model that combines a first random effect associated with the markers and a second effect called polygenic effect that considers the relationship matrix of genotyped animals. According to Muir et al (28), the phenotype classical selection (BLUP) can capture QTL regardless of if there is a linkage disequilibrium between QTL and the markers, meaning that it does not use molecular markers. This favors the model used in the bayes C method, since it can capture the genetic variance of QTL's in the linkage disequilibrium with the evaluated markers and the remaining genetic variance through the polygenic effect.

Moser et al (11) used 7372 SNP genotypes and EBV's (average reliability 89%) from 1945 Holstein bulls to predict the molecular estimated breeding value (MEBV) in PGRA, using the bayes R method. The precision and the bias of the prediction of the MEBV were calculated through a cross validation using a training group of 706 young bulls in which the MEBV was estimated and compared to the conventional EBV.

The correlation between the MEBV and the conventional EBV for PGRA was moderate ( $r_{p_{MEBV;EBV}}=0.56$ ). Notwithstanding, when MEBV is combined with the polygenic effect obtained from the pedigree to obtain the GEBV, the correlation coefficient was higher ( $r_{p_{GEBV;EBV}}=0.62$ ). These results corroborated the results obtained in this research for PGRA ( $r_{MEBV;EBV}=0.967$ ;  $r_{GEBV;EBV}=0.995$ ), indicating the importance of including the polygenic effect in genomic evaluations since GEBV considers a portion of the genetic variance captured for the markers and the polygenic variance not captured by the markers. However, in both investigations, the correlations are different, being higher in the investigation herein. The explanation for this difference in the results is because in this investigation, the genomic values (MEBV and GEBV) were estimated in a reference population, while in Moser et al (11) investigation, the genomic values were estimated in a validation population.

The difference between the assumptions used for each estimation method must also be considered. Bayes C assumes a fraction ( $\pi$ ) of SNP's without

convencional). Según Muir et al (28), la selección fenotípica clásica (BLUP) puede capturar QTL independientemente de que haya desequilibrio de ligamiento entre QTL y marcadores, es decir que no necesita del uso de marcadores moleculares. Esto favorece el modelo empleado en el método bayes C, ya que puede capturar la varianza genética de los QTL en desequilibrio de ligamiento con los marcadores evaluados y la varianza genética restante a través del efecto poligénico.

Moser et al (11) usaron los genotipos de 7372 SNPs y los EBVs (confiabilidad promedio 89%) de 1945 toros Holstein para predecir el valor de cría molecular (MEBV) en PGRA, usando el método bayes R. La precisión y el sesgo de predicción del MEBV fue calculada mediante validación cruzada usando un grupo de entrenamiento de 706 toros jóvenes en los cuales se estimó el MEBV, que fue comparado con el EBV convencional.

La correlación entre el MEBV y el EBV convencional para PGRA fue moderada ( $r_{p_{MEBV;EBV}} = 0.56$ ). No obstante, cuando se combina el MEBV con el efecto poligénico obtenido del pedigrí para obtener el GEBV, el coeficiente de correlación fue mayor ( $r_{p_{GEBV;EBV}} = 0.62$ ). Estos resultados corroboran los obtenidos en la presente investigación para la característica PGRA ( $r_{MEBV;EBV} = 0.967$ ;  $r_{GEBV;EBV} = 0.995$ ), indicando la importancia de incluir el efecto poligénico en las evaluaciones genómicas, dado que el GEBV considera la proporción de la varianza genética capturada por los marcadores y la varianza poligénica no capturada por estos. Sin embargo, en las dos investigaciones, las correlaciones son diferentes, siendo mayores en la actual investigación. La explicación de esta diferencia en los resultados, se debe a que en la presente investigación, los valores genómicos (MEBV y GEBV) fueron estimados en una población de referencia, mientras que en la investigación de Moser et al (11), fueron estimados en una población de validación.

También debe considerarse la diferencia en los supuestos que usa cada método de estimación. Bayes C asume una fracción ( $\pi$ ) de SNPs sin efecto ni varianza y la fracción restante ( $1 - \pi$ ) con efectos, los cuales se distribuyen como una normal, mientras que en Bayes R, aunque es una extensión de bayes C; los efectos de los SNPs se asumen con múltiples distribuciones normales (12).

De igual manera, Moser et al (11), determinaron el coeficiente de regresión del EBV sobre el MEBV ( $b_{MEBV;EBV}$ ) para PPRO y obtuvieron un valor 1.06 para el método de estimación bayes R. En la presente investigación, el coeficiente de regresión

effect or variance and the remaining fraction (1-n) with effects, which are distributed as a normal fraction; while in Bayes R, although an extension of bayes C, the effects of SNP's are assumed as multiple normal distributions (12).

Likewise, Moser et al (11), determined the regression coefficient of the EBV on the MEBV ( $b_{\text{MEBV};\text{EBV}}$ ) for PPRO and they obtained a value of 1.06 for the bayes R estimation method. In this investigation, the regression coefficient ( $b_{\text{MEBV};\text{EBV}}$ ) was 1.015, very similar to that obtained by the mentioned authors. These results indicate that the change in magnitude of the MEBV's with respect to conventional EBV's is very small. A value greater than one suggests that the markers do not capture the whole genetic variance, while a value less than 1 indicates that the genomic prediction would be inflated with respect to the conventional method (BLUP) and if such is the case, the genomic values would be biased with respect to the conventional EBV's. On the other hand, the regression coefficient of EBV on GEBV for PGRA was slightly less than 1 ( $b_{\text{GEBV};\text{EBV}} = 0.988$ ), because when combining the effects of the markers and the polygenic effect, the GEBV captures a greater portion of the genetic variance and therefore, it will be closer than 1 compared to the MEBV.

Verbyla et al (23), in a simulation study, evaluated 4 bayesian methods; Bayes BLUP, Bayes A, Bayes A/B and Bayes C, which were used to estimate GEBVs. GEBVs were compared to true estimated breeding value (TEBV), which were obtained by simulation. The correlation results between TEBV and GEBV, ( $r_{\text{GEBV};\text{TEBV}}$ ) were 0.885, 0.857, 0.889 and 0.861 and the regression coefficients ( $b_{\text{GEBV};\text{TEBV}}$ ) were: 0.979, 1.162, 1.081 and 1.024 respectively for each method. These results show that the use of relevant a priori information in bayesian regressions is of great importance in the estimation of genomic values, allowing obtaining more precise and less biased predictions.

In this investigation, correlations ( $r_{\text{pGEBV};\text{EBV}}$ ) greater than 0.98 were obtained for PL, PPRO, PGRA and SCS and regression coefficients ( $b_{\text{GEBV};\text{EBV}}$ ) that range between 0.979 and 1.008 for the same traits. However, although the correlations between MEBV's and conventional EBV's were slightly lower in all of the evaluated characteristics ( $r_{\text{pMEBV};\text{EBV}}$  greater than 0.93). These results are similar to those obtained by Verbyla et al (23), whom suggest that the bayesian methods are consistent in simulated data and real data.

Verbyla et al (23), also determined that the correlation by ranges for 100 animals and

( $b_{\text{MEBV};\text{EBV}}$ ) fue de 1.015, muy similar al obtenido por los autores mencionados. Estos resultados indican que el cambio en magnitud de los MEBVs con respecto a los EBVs convencionales es muy pequeño. Un valor mayor de uno sugiere que los marcadores no capturaron toda la varianza genética, mientras que un valor menor de 1 indica que la predicción genómica sería inflada con respecto al método convencional (BLUP) y de ser así, los valores genómicos serían sesgados con respecto a los EBVs convencionales. Por otra parte, el coeficiente de regresión del EBV sobre el GEBV para PGRA fue levemente menor de 1 ( $b_{\text{GEBV};\text{EBV}} = 0.988$ ), dado que al combinar los efectos de los marcadores y el efecto poligénico, el GEBV captura mayor proporción de la varianza genética y por tanto será más cercano a 1 comparado con el MEBV.

Verbyla et al (23), en un estudio de simulación, evaluaron 4 métodos bayesianos; Bayes BLUP, Bayes A, Bayes A/B y Bayes C, los cuales fueron empleados para estimar GEBVs. Los GEBVs fueron comparados con los valores genéticos verdaderos (TEBV) que fueron obtenidos por simulación. Los resultados de correlación entre TEBV y GEBV, ( $r_{\text{GEBV};\text{TEBV}}$ ) fueron 0.885, 0.857, 0.889 y 0.861 y Los coeficientes de regresión ( $b_{\text{GEBV};\text{TEBV}}$ ) fueron: 0.979, 1.162, 1.081 y 1.024 respectivamente para cada método. Estos resultados muestran ver que el uso de una información a priori relevante en regresiones bayesianas es de gran importancia en la estimación de valores genómicos, permitiendo obtener predicciones más precisas y menos sesgadas.

En la presente investigación, se obtuvieron correlaciones ( $r_{\text{pGEBV};\text{EBV}}$ ) mayores de 0.98 para PL, PPRO, PGRA y SCS y coeficientes de regresión ( $b_{\text{GEBV};\text{EBV}}$ ) que oscilaron entre 0.979 y 1.008 para las mismas características. Sin embargo, aunque las correlaciones entre los MEBV y los EBVs convencionales fueron ligeramente más bajas en todas las características evaluadas ( $r_{\text{pMEBV};\text{EBV}}$  mayores de 0.93). Estos resultados son similares que los obtenidos por Verbyla et al (23), quienes sugieren que los métodos bayesianos son consistentes en datos simulados y datos reales.

Verbyla et al (23), también determinaron la correlación por rangos para 100 animales y obtuvieron valores altos (0.691, 0.696, 0.73 y 0.71) para cada método respectivamente. En la presente investigación también fue determinada la correlación de Spearman ( $r_{\text{sGEBV};\text{EBV}}$ ). Los valores de correlación para PL, PPRO, PGRA y PL fueron: 0.991, 0.983, 0.989 y 0.986 y se consideran altos, lo que indica que no hay cambios significativos en la clasificación o ranking de los animales. Las correlaciones por rango de Spearman entre el

obtained high values (0.691, 0.696, 0.73 and 0.71) for each method, respectively. In this investigation a Spearman correlation ( $r_{s_{\text{GEBV,EBV}}}$ ) was also determined. The correlation values for PL, PPRO, PGRA and PL were: 0.991, 0.983, 0.989 and 0.986 and they are considered high, which indicates that there are no significant changes in the classification or ranking of animals. Spearman correlation by ranges between the MEBV and the EBV ( $r_{s_{\text{MEBV,EBV}}}$ ) were also high (greater than 0.93), but slightly less than those obtained between the GEBV and the EBV.

In another study, Legarra et al (30) used 51325 SNPs and 2DYDs (double production of daughters) from 1756 Holstein bulls to predict genomic values (GEBV) for PL, PPRO and PGRA. The effects of the markers were estimated using the Bayes LASSO method (BL2Var). The estimation of the GEBVs was conducted through cross validation using a training group of 1216 young bulls and was compared with the 2DYD. The correlations ( $r_{\text{GEBV;2DYD}}$ ) for PL, PGRA and PPRO were: 0.41, 0.73 y 0.48 and the regression coefficients ( $b_{\text{GEBV;2DYD}}$ ) were: 0.67, 1.18 and 1.10 respectively. The results showed that the correlations ( $r_{\text{GEBV;2DYD}}$ ) were higher in PPRO and PGRA compared to the production of milk. The authors suggest that a regression coefficient less than 1 ( $b < 1$ ), indicates an inflation in the genetic variance in milk production traits like in the case of PL; while in the milk quality traits, the largest part of the genetic variation is captured by large QTL effects; therefore, the regression coefficients tend to be greater than 1 ( $b > 1$ ), like in the case of PPRO and PGRA. In this investigation it can be considered a similar behavior in PGRA, in which a regression coefficient greater than 1 ( $b_{\text{MEBV;EBV}} = 1.015$ ) was obtained, while in SCS it was less than 1 ( $b_{\text{MEBV;EBV}} = 0.95$ ).

In conclusion, the predictions of genomic values (GEBV) and molecular values (MEBV) using bayes C method were consistent with the predictions of the EBVs estimate through the conventional method (BLUP); allowing to obtain high associations between all evaluated traits and a high association in the ranking or classification of the animals established by their genomic values (MEBV and GEBV) and conventional EBV's. On the other hand, considering the regression coefficient ( $b_{\text{GEBV;EBV}}$ ), a significant change in the magnitude between the genomic values (MEBV and GEBV) and conventional EBVs did not occur, given that the regression coefficients were close to one.

However, in this investigation it must be considered that the effects of the markers were estimated in a small reference population

MEBV y el EBV ( $r_{s_{\text{MEBV,EBV}}}$ ) también fueron altas (mayores de 0.93), pero levemente inferiores que las obtenidos entre el GEBV y el EBV.

En otro estudio, Legarra et al (30), usaron los genotipos de 51325 SNPs y los 2DYDs (doble producción de las hijas) de 1756 toros Holstein para predecir valores genómicos (GEBV) para PL, PPRO y PGRA. Los efectos de los marcadores fueron estimados usando el método Bayes LASSO (BL2Var). La estimación de los GEBVs fue realizada mediante validación cruzada usando un grupo de entrenamiento de 1216 toros jóvenes y fueron comparados con el 2DYD. Las correlaciones ( $r_{\text{GEBV;2DYD}}$ ) para PL, PGRA y PPRO fueron: 0.41, 0.73 y 0.48 y los coeficientes de regresión ( $b_{\text{GEBV;2DYD}}$ ) fueron: 0.67, 1.18 y 1.10 respectivamente. Los resultados mostraron que las correlaciones ( $r_{\text{GEBV;2DYD}}$ ) fueron más altas en PPRO y PGRA comparadas con producción de leche. Los autores sugieren que un coeficiente de regresión menor de 1 ( $b < 1$ ), indica una inflación de la varianza genética en características de producción como es el caso de PL, mientras que en características composicionales, la mayor parte de la variación genética es capturada por grandes efectos QTL, por lo cual, los coeficientes de regresión tienden a ser mayores de 1 ( $b > 1$ ), como es el caso de PPRO y PGRA. En la presente investigación puede considerarse un comportamiento similar en PGRA, en la cual se obtuvo un coeficiente de regresión mayores que 1 ( $b_{\text{MEBV;EBV}} = 1.015$ ), mientras que en SCS fue menor de 1 ( $b_{\text{MEBV;EBV}} = 0.95$ ).

En conclusión, se considera que la estimación de valores genómicos (GEBV) y valores moleculares (MEBV) usando el método Bayes C fue consistente con los EBVs convencionales estimados mediante el método BLUP; permitiendo obtener altas asociaciones entre los valores genéticos en todas las características evaluadas y una alta asociación en el ranking o clasificación de los animales establecida por sus valores genómicos (MEBV y GEBV) y EBVs convencionales. Por su parte, teniendo en cuenta los coeficiente de regresión ( $b_{\text{GEBV;EBV}}$ ), no se dió un cambio significativo en la magnitud entre los valores genómicos (MEBV y GEBV) y los EBVs convencionales, dado que los coeficientes de regresión fueron muy cercanos a uno.

Sin embargo, debe considerarse que en la presente investigación fueron estimados los efectos de los marcadores en una población de referencia pequeña, como punto de partida de la implementación de selección genómica y se trata de abordar la metodología presentando algunos resultados iniciales de comparación entre un modelo de selección genómica y

as a starting point in the implementation for the genomic selection and the methodology addressed presents some initial comparison results between the genomic selection model and the conventional model. Notwithstanding, it is indispensable to increase the number of genotyped animals and the number of SNPs in order to consolidate a reference population on which more reliable genomic evaluations can be conducted.

el modelo convencional. No obstante, es indispensable incrementar el número de animales genotificados y el número de SNPs para lograr consolidar una población referencia de la cual se pueda realizar evaluaciones genómicas más confiables.

## REFERENCES

- Henderson CR. Applications of linear models in animal breeding. Guelph: CGIL Publications; 1984.
- Fisher R. The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh 1918; 52:399-433.
- Cole JB, VanRaden PM, O'Connell JRO, Van Tassell CP, Sonstegard TS, Schnabel RD et al. Distribution and location of genetic effect for dairy traits. J Dairy Sci 2009; 92(6):2931-2946.
- Kahi AK, Rewe TO, Kosgey IS. Sustainable community-based organizations for the genetic improvement of livestock in developing countries. Outlook Agric 2005; 34(4):261-270.
- Godard ME, Hayes BJ. Genomic Selection. J Anim Breed Genet 2007; 124(6):323-330.
- Meuwissen THE, Hayes B, Goddard M. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001; 157(4):1819-1829.
- Dekkers JC. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. J Anim Sci 2004; 82(E-Suppl):E313-328.
- Hayes BJ, Bowman PJ, Chamberlain AC, Goddard ME. Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 2009; 92(2):433-443.
- Schepers J, Wigal KA. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim Front 2012; 12(1):4-9.
- Calus MPL. Genomic breeding value prediction: methods and procedures. Animal 2010; 4(2):157-164.
- 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.
- Meuwissen T, Hayes B, Goddard M. Accelerating Improvement of livestock with Genomic Selection. Annu Rev Anim Biosci 2013; 1:221-237.
- VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 2009; 92 (1):16-24.
- Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM et al. Genomic selection in the French Lacaune dairy sheep breed. J Dairy Sci 2012; 95(5):2723-2733.
- 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.
- Ali AK, Shook GE. An Optimun transformation for somatic cell concentration in milk. J Dairy Sci 1980; 63(3): 487-490.
- Echeverri J, López A, Parra J. Software control 1. [CD-ROM]. Versión 1. Medellín: Universidad Nacional de Colombia sede Medellín; 2010.
- SAS. Statistical Analysis Systems [CD-ROM]. Versión 9.1 Cary, NC, USA: SAS Inst, Inc; 2006.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007; 81(3):559-575.

20. Mrode RA, Thompson R. Linear models for the prediction of animal breeding values. Cambridge: CABI Publishing; 2005.
21. Boldman K, Kriese L, Van Vleck L, Van Tassell C, Kachman S. MTDFREML: A Set of programs to obtain estimates of variances and covariances. [Programa de computadora]. Clay Center (NE): USDA-ARS; 1995.
22. Kizilkaya k, Fernando RL, Garrick DJ. Genomic Prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci* 2010; 88(2):544-551.
23. Verbyla KL, Bowman PJ, Hayes BJ, Raadsma H, Goddard ME. Sensitivity of genomic selection to using different prior distributions. *BMC Proc* 2010; 4(1):S5.
24. Legarra A, Ricard A, Filangi O. GS3: Genomic selection, Gibbs sampling, Gauss Seidel and Bayes Cpi. [Programa de computadora] Toulouse: Inra; 2013.
25. Mäntysaari E, Zengting L, Van Raden P. Interbull Valdation Test for Genomics Evaluations. *Interbull Bolletin* 2010; 41:17-22.
26. Lillehammer L, Meuwissen THE, Sonesson AK. A comparison of dairy cattle breeding designs that use genomic selection. *J Dairy Sci* 2010; 94:493-500.
27. Calus MPL, Veerkamp RF. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *J Anim Breed Genet* 2007; 124:362-368.
28. Muir WM. Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet* 2007; 124:342-355.
29. Visscher PM, Yang J, Goddard MEA. A commentary on "common SNPs explain a large proportion of the heritability for human height" by Yang et al. (2010). *Twin Res Hum Genet* 2012; 13:517-524.
30. Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection. *Genet Res Camb* 2011; 93(1):77-87.