

COMPARACIÓN DE LA EXACTITUD DE VALORES GENÓMICOS DE ANIMALES PREDICHOS A TRAVÉS DEL ANÁLISIS CON DOS MODELOS ALTERNATIVOS

COMPARISON OF THE ACCURACY OF GENOMIC VALUES OF ANIMALS PREDICTED THROUGH ANALYSIS WITH TWO ALTERNATIVE MODELS

Baldomero Alarcón-Zúñiga, Fernanda Ramírez-Flores, Agustín Ruíz-Flores*, Rodolfo Ramírez-Valverde, Luis A. Saavedra-Jiménez, José L. Zepeda-Batista

Posgrado en Producción Animal. Departamento de Zootecnia. Universidad Autónoma Chapingo. 56230. Chapingo, Estado de México. (arf@correo.chapingo.mx)

RESUMEN

El valor genómico es la mejor predicción del valor genotípico de un animal, cuya exactitud varía en función de varios factores. Los objetivos de este estudio fueron, mediante simulación, comparar la exactitud de los valores genómicos de animales predichos a través del análisis con dos modelos alternativos, y obtener la correlación genética entre estos y los valores genéticos verdaderos simulados. Una población con tamaño efectivo de 800 individuos fue simulada y se usaron 100 generaciones para generar desequilibrio de ligamiento. Después se simuló otra población con 14 generaciones, un panel de 53 010 polimorfismos de nucleótido simple (SNPs) ubicados aleatoriamente en 30 cromosomas y 540 *loci* de características cuantitativas. También se simularon los genotipos y fenotipos de 6400 animales usando una heredabilidad=0.40 y considerando solamente los efectos aditivos. Cuatro conjuntos de valores moleculares fueron entrenados con los marcadores SNP's de las generaciones 7 a 10, y los correspondientes valores genéticos ajustados. Cuando la población de entrenamiento (PEn) tuvo tamaño mayor, los estimadores de exactitud (R^2) fueron mayores. Similarmente, cuando la población de evaluación (PEv) estuvo más cerca a la PEn, los estimadores de R^2 fueron mayores y se fueron reduciendo cuando la distancia entre PEn y PEv se incrementaba. El estimador de la varianza del error de predicción (PEV) fue menor (0.15 ± 0.01) cuando PEn y PEv estuvieron más cerca, independientemente del tamaño de PEn, generación 10 para PEn y generación 11 para PEv. Por el contrario, el PEV mayor (0.29 ± 0.02) se obtuvo cuando la PEn incluyó los animales de la generación siete y se evaluó la generación 14; es decir, cuando la distancia entre PEn y PEv era mayor.

ABSTRACT

The genomic value is the best prediction of the genotypic value of an animal, whose accuracy varies in function of several factors. The objectives of this study were to compare, through simulation, the accuracy of the genomic values of animals predicted through analysis with two alternative models, and to obtain the genetic correlation between these and the true genetic values simulated. A population with an effective size of 800 individuals was simulated and 100 generations were used to generate linkage disequilibrium. Then, another population was simulated with 14 generations, a panel of 53 010 single nucleotide polymorphisms (SNPs), placed randomly in 30 chromosomes and 540 quantitative traits loci. The genotypes and phenotypes of 6400 animals were also simulated using a heritability=0.40, and considering only the additive effects. Four sets of molecular values were trained using the SNP markers of generations 7 to 10, and the corresponding adjusted genetic values. When the training population (TP) was larger in size, the accuracy estimators (R^2) were higher. Similarly, when the evaluation population (EP) was closer to the TP, the R^2 estimators were higher and they became smaller when the distance between TP and EP increased. The estimate of the prediction error variance (PEV) was lower (0.15 ± 0.01) when TP and EP were closer, regardless of the size of the TP, generation 10 for TP and generation 11 for EP. On the contrary, the highest PEV (0.29 ± 0.02) was obtained when the TP included animals from generation seven and generation 14 was evaluated; that is, when the distance between TP and EP was greater.

Keywords: Predicted genomic value, accuracy of the genomic value, genomic selection.

* Autor responsable ♦ Author for correspondence.

Recibido: marzo, 2015. Aprobado: junio, 2015.

Publicado como ARTÍCULO en *Agrociencia* 49: 613-622. 2015.

Palabras clave: Valor genómico predicho, exactitud del valor genómico, selección genómica.

INTRODUCCIÓN

Un objetivo común en programas de mejoramiento genético es usar la información del ADN para aumentar el progreso genético al reducir el intervalo entre generaciones y aumentar la exactitud de las predicciones. Ahora se puede determinar el genotipo de animales para decenas de miles de polimorfismos de nucleótido simple (SNP's, siglas en inglés; Goddard y Hayes, 2007). Estos SNP's pueden usarse para obtener los valores genómicos de los animales (Meuwissen *et al.*, 2001).

La evaluación genómica de animales puede efectuarse en una (Miszta *et al.*, 2009; Aguilar *et al.*, 2011), o en dos etapas (Hayes *et al.*, 2009; Van Raden *et al.*, 2009). Cuando se realiza en dos etapas, en la primera se usa una población de entrenamiento (PEn) para obtener las ecuaciones de predicción, las cuales se usarán en una segunda etapa en la población de evaluación (PEv) para predecir los valores genómicos (VGn); en la PEv no se requiere tener los registros fenotípicos de los individuos, pero sí sus genotipos. Los animales utilizados en la PEv no deberán considerarse en la PEn (Garrick, 2011). En la fase de entrenamiento se predice la influencia de pequeñas regiones genómicas por regresión de la información observada del genotipo de los marcadores para una población dada. El genotipo deberá estar constituido por un gran número de marcadores (Garrick *et al.*, 2009). Los animales en ambas poblaciones deberán ser de la misma raza (Kachman *et al.*, 2013), y pueden evaluarse al obtener una muestra de su tejido para realizar el genotipado, lo cual permite reducir el intervalo entre generaciones y, como consecuencia, aumentar el progreso genético.

La exactitud de los valores genómicos predichos se define como la correlación entre el VGn y el valor genético verdadero (VG) (Garrick y Saatchi, 2011; Meuwissen *et al.*, 2001). Un incremento en la exactitud de los VGn de los animales candidatos a selección incrementará el progreso genético, especialmente para características con baja heredabilidad (h^2).

Habier *et al.* (2007) indicaron que tanto el desequilibrio de ligamiento entre los marcadores y los *loci* de características cuantitativas (QTL, por sus siglas en inglés) y las relaciones de parentesco entre

INTRODUCTION

A common objective in the genetic improvement programs is to use the DNA information to increase the genetic progress when reducing the interval between generations and increasing the accuracy of the predictions. It is now possible to determine the genotype of animals for tens of thousands of single nucleotide polymorphisms (SNPs) (Goddard and Hayes, 2007). These SNPs can be used to obtain the genomic values of the animals (Meuwissen *et al.*, 2001).

The genomic evaluation of animals can be done in one (Miszta *et al.*, 2009; Aguilar *et al.*, 2011), or two stages (Hayes *et al.*, 2009; Van Raden *et al.*, 2009). When it is done in two stages, a training population (TP) is used in the first one, to obtain the prediction equations, which will be used in the second stage with the evaluation population (EP) to predict the genomic values (GV); in the EP there is no need for phenotypic records of individuals, but their genotypes are required. The animals used in the EP should not be considered in the TP (Garrick, 2011). During the training phase the influence of small genomic regions is predicted through regression of the information observed from the markers' genotypes for a given population. The genotype should be constituted by a large number of markers (Garrick *et al.*, 2009). The animals in both populations should be of the same breed (Kachman *et al.*, 2013), and can be evaluated when obtaining a sample of their tissue to perform the genotyping, which allows reducing the generation interval and, as consequence, increasing the genetic progress.

The accuracy of the genomic values predicted is defined as the correlation between the GV and the true genetic value (TGV) (Garrick and Saatchi, 2011; Meuwissen *et al.*, 2001). An increase in the accuracy of the GV of the animals that are candidates for selection would increase the genetic progress, especially for low heritability traits (h^2).

Habier *et al.* (2007) indicated that both the linkage disequilibrium between markers and the quantitative trait *loci* (QTL), and the relationships between the TP and EP individuals, influence the accuracy of the GV. Close relationships between individuals from the TP and EP, yield better predictions than those of individuals with more distant relationships (Habier *et al.*, 2010; Saatchi *et al.*, 2010; Saatchi *et al.*, 2011).

individuos de las PEn y PEv, influyen en la exactitud de los VGn. Relaciones cercanas entre los individuos de PEn y PEv tienen mejores predicciones que las de individuos con relaciones más distantes (Habier *et al.*, 2010; Saatchi *et al.*, 2010; Saatchi *et al.*, 2011). Otros factores que también influyen en la exactitud de los VGn son la distancia entre marcadores, tamaño de la PEn, número de animales o sementales en la PEn, heredabilidad de la característica, estructura genética, y número de marcadores usados para la predicción genómica (Goddard y Hayes, 1992; Meuwissen *et al.*, 2001; Hayes *et al.*, 2009).

Con información ajustada, de acuerdo con Garrick *et al.* (2009), se puede obtener una exactitud hasta 2.76 veces mayor que con el registro de un solo individuo. Ostensen *et al.* (2011) obtuvieron exactitudes 18 a 39 % mayores de los VGn, dependiendo de la característica evaluada, cuando utilizaron VG ajustados por los valores genéticos de los padres del animal (DEVG, siglas en inglés) en lugar de VG, como variable de respuesta. Garrick *et al.* (2009) mostraron que los DEVG con el promedio parental removido producen VGn predichos más exactos, por dos razones: 1) los DEVG como variable respuesta resultan en menos conteos dobles que con los VG porque los DEVG excluyen información de los ancestros, y si tanto la progenie como sus padres son genotipados, el grado de conteo doble decrece cuando se usan DEVG como la variable de respuesta; 2) al usar los VG como las variables de respuesta el grado de doble conteo en los VGn decrece, en particular cuando las confiabilidades de los valores genéticos son bajas.

La predicción del VGn para un animal utilizando un enfoque Bayesiano, como el Bayes C (Meuwissen *et al.*, 2001; Habier *et al.*, 2011), es la media *a posteriori* del valor genético del animal. La exactitud puede obtenerse elevando al cuadrado la correlación entre el VG verdadero y los VGn.

Los objetivos de este estudio fueron comparar la exactitud de los valores genómicos predichos con dos modelos de análisis utilizando datos simulados, y estimar la correlación genética entre los valores genéticos verdaderos obtenidos mediante simulación con los valores genómicos predichos con el programa Gen-Sel.

Other factors that also influence the accuracy of the GV are the distance between markers, the size of the EP, the number of animals or studs in the TP, the heritability of the trait, the genetic structure, and the number of markers used for the genomic prediction (Goddard and Hayes, 1992; Meuwissen *et al.*, 2001; Hayes *et al.*, 2009).

With adjusted information, according to Garrick *et al.* (2009), an accuracy of up to 2.76 times higher than with the record of a single individual can be obtained. Ostensen *et al.* (2011) obtained accuracies 18 to 39 % higher than the GV, depending on the trait evaluated, when deregressed GV (DEVG) were used, instead of TGVs as a response variable. Garrick *et al.* (2009) showed that the DEVGs with the removed parent average produce more exact VGs, because of two reasons: 1) the DEVGs as response variable result in less double counts than with the TGVs because the DEVGs exclude information from the ancestors, and if both the offspring and its parents are genotyped, the degree of double count decreases when DEVGs are used as the response variable; 2) when using the TGVs as the response variable, the degree of double count in the GVs decreases, particularly when the reliabilities of the genetic values are low.

The prediction of the GV for an animal using a Bayesian approach, such as Bayes C (Meuwissen *et al.*, 2001; Habier *et al.*, 2011), is the *a posteriori* mean of the genetic value of the animal. The accuracy can be obtained squaring the correlation between the TGV and the GVs.

The objectives of this study were to compare the accuracy of the genomic values predicted with two analysis models using simulated data, and to estimate the genetic correlation between the true genetic values obtained through simulation and the genomic values predicted with the Gen-Sel program.

MATERIALS AND METHODS

The linkage disequilibrium was obtained by simulating a historical population with an effective size of 800 individuals and 100 generations. Then, another population with 14 discrete generations was simulated, 20 males and 200 females selected randomly to produce the following generation, a panel of 53 010 SNPs located randomly in 30 chromosomes,

MATERIALES Y MÉTODOS

El desequilibrio de ligamiento se obtuvo simulando una población histórica con tamaño efectivo de 800 individuos y 100 generaciones. Después se simuló otra población con 14 generaciones discretas, 20 machos y 200 hembras seleccionadas aleatoriamente para producir la siguiente generación, un panel de 53 010 SNP's ubicados aleatoriamente en 30 cromosomas, cada uno separado 100 centiMorgans de longitud, y 540 QTL con efectos provenientes de una distribución gama (Weller *et al.*, 2005). Los SNP's y los QTL se consideraron bialélicos con frecuencias iniciales aleatorias. También se simularon los genotipos y fenotipos de 6400 bovinos usando una heredabilidad de 0.4, y sólo se consideraron los efectos aditivos. Otros parámetros de la población simulada se indican en el Cuadro 1. Para generar la información genotípica y fenotípica se utilizó el programa QMSim (Sargolzaei y Schenkel, 2009).

Cuatro conjuntos de valores moleculares fueron entrenados usando los marcadores SNP's de las generaciones 10 (n=1,000), 9 y 10 (n=1,400), 8 a 10 (n=1800), 7 a 10 (n=2200), y los correspondientes DEVG. Posteriormente se predijeron los valores genéticos con un modelo mixto que incluyó los efectos aleatorio del animal y fijos de sexo del animal y generación mediante el programa ASReml (Gilmour *et al.*, 2009) y los DEVGs se obtuvieron siguiendo la metodología de Garrick *et al.* (2009). La ponderación para el *i*-ésimo animal se obtuvo con la ecuación $w_i = (1 - h^2) / [(c + (1 - r^2)/r^2)h^2]$, donde *c* es la falta de ajuste de la ecuación de predicción o la parte genética no explicada por los marcadores, se asumieron valores de *c*=0.1 y $h^2=0.4$; y r^2 fue la confiabilidad de los DEVG para el *i*-ésimo animal. El modelo usado para obtener los efectos de los marcadores en la población de entrenamiento fue:

$$y_i = \mu + \sum_{j=1}^k z_{ij} u_j + e_i$$

donde y_i es el DEVG para el *i*-ésimo animal, μ es la media poblacional, *k* es el número de *loci* de los marcadores en el panel, z_{ij} es el genotipo del *i*-ésimo animal (número de copias del alelo 1) en el *j*-ésimo *locus* del marcador, u_j es el efecto aleatorio del *j*-ésimo SNP muestreado de $u_j \sim N(0, \sigma_u^2)$ con probabilidad $1 - \pi$ o 0 con probabilidad π , y e_i es el residual con varianza σ_e^2/w_i , dependiendo de la exactitud del DEVG para el *i*-ésimo individuo. Los valores *a priori* para σ_a^2 y σ_e^2 fueron 0.4 y 0.04, respectivamente.

Las ecuaciones de predicción genómica para obtener los valores genómicos se derivaron de un análisis ponderado para los animales de las generaciones 11 a 14, para lo cual se usó la función BayesCp del programa Gen-Sel (Fernando y Garrick, 2009)

each one separated by 100 centiMorgans of length, and 540 QTLs with effects from a gamma distribution (Weller *et al.*, 2005). Both the SNPs and the QTLs were considered to be biallelic with random initial frequencies. The genotypes and phenotypes of 6400 bovinos were simulated using a heritability of 0.4, and only the additive effects were considered. Other parameters of the population simulated are shown in Table 1. In order to generate the genotypic and phenotypic information, the QMSim software was used (Sargolzaei and Schenkel, 2009).

Four sets of molecular values were trained using the SNP markers of generations 10 (n=1000), 9 and 10 (n=1400), 8 to 10 (n=1800), 7 to 10 (n=2200), and the corresponding DEVGs. Later, the genetic values were predicted with a mixed model that included the random effects of the animal and the fixed ones of animal sex and generation through the ASReml software (Gilmour *et al.*, 2009), and the DEVGs were obtained by following the methodology by Garrick *et al.* (2009). Weighting for the *i*-th animal was obtained with the equation $w_i = (1 - h^2) / [(c + (1 - r^2)/r^2)h^2]$, where *c* is the lack of fit of the prediction equation or the genetic part not explained by the markers, values of *c*=0.1 and $h^2=0.4$ were assumed; and r^2 was the reliability of the DEVGs for the *i*-th animal. The model used to obtain the effects from the markers in the training population was:

Cuadro 1. Estructura poblacional y parámetros utilizados para la simulación.

Table 1. Population structure and parameters used for the simulation.

Parámetro	Valor
Número de cromosomas	30
Número de SNP's	53 010
Número de alelos marcadores por locus	2
Frecuencias iniciales de los alelos marcadores	Aleatorias
Número de QTL	540
Número de alelos para los QTL	2
Frecuencias iniciales de los alelos QTL	Aleatorias
Número de generaciones para crear desequilibrio de ligamiento	100
Número y tamaño de generaciones	1 a 9400 animales 10, 1000 animales 11 a 14, 400 animales
Generaciones usadas como población de entrenamiento	7 a 10
Generaciones evaluadas	11 a 14
Heredabilidad	0.4

a través de la plataforma bioinformática para implementar selección genómica (Bioinformatics to Implement Genomic Selection, BIGS) de la Universidad Estatal de Iowa (EE.UU.)¹. La opción BayesCp se seleccionó por sus ventajas sobre otras opciones de Bayes, al ajustar un modelo similar a BayesB (Meuwissen *et al.*, 2001; Habier *et al.*, 2011), además de estimar la proporción de *loci* no informativos en el modelo (π), y el supuesto de una varianza común para todos los *loci* con efectos no cero, en lugar de la varianza individual para cada *locus* (Habier *et al.*, 2011). La cadena usada tenía una longitud de 41,000, se utilizaron las últimas 40 000 muestras para obtener los estimadores *a posteriori* de los efectos medios y varianzas, se quemaron las primeras 1000 iteraciones, y p se fijó en 0.95.

El valor genómico para el *i*-ésimo animal se obtuvo como la suma sobre todos los marcadores de las medias *a posteriori* de los efectos predichos de los SNP's, multiplicados por el número de copias del alelo 1, usando la ecuación:

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

donde VGn_i es el valor genético molecular, valor genómico, para el *i*-ésimo animal, z_{ij} es el genotipo del *j*-ésimo marcador en el *i*-ésimo animal, y $\hat{u}_j$ es la media *a posteriori* del efecto del SNP para el *j*-ésimo marcador.

Los modelos utilizados para obtener los valores genómicos en Gen-Sel fueron: 1) uno sin considerar los efectos de sexo y número de generación, y 2) otro que sí los incluyó. Los criterios de comparación utilizados fueron la exactitud, estimada a través del coeficiente de determinación múltiple (R^2) y la varianza del error de predicción de los valores genómicos (PEV). La correlación genética entre los valores genéticos verdaderos simulados obtenidos con el programa QMSim (Sargolzaei y Schenkel, 2009) y los valores genómicos predichos con el programa Gen-Sel (Fernando y Garrick 2009) se estimó con el programa ASReml (Gilmour *et al.*, 2009).

RESULTADOS Y DISCUSIÓN

Los estimadores de R^2 fueron mayores cuando PEn fue mayor (incluyó más generaciones) que PEv. De manera similar, los estimadores de R^2 fueron mayores cuando disminuía la distancia entre PEv y Pen, y se fueron reduciendo cuando aumentaba la distancia entre ambas poblaciones (Cuadro 2).

$$y_i = \mu + \sum_{j=1}^k z_{ij} u_j + e_i$$

where y_i is the DEVG for the *i*-th animal, μ is the population mean, k is the number of *loci* of the markers in the panel, z_{ij} is the genotype of the *i*-th animal (number of copies of allele 1) in the *j*-th *locus* of the marker, u_j is the random effect of the *j*-th SNP sampled from $u_j \sim N(0, s_u^2)$ with a probability of $1-p$ or 0 with probability p , and e_i is the residual with variance s_e^2/w_p , depending on the accuracy of the DEVG for the *i*-th individual.

The *a priori* values for s_a^2 and s_e^2 were 0.4 and 0.04, respectively. The equations for genomic prediction used to obtain genomic values were derived from an analysis weighted for animals of generations 11 to 14, for which the BayesCp function of the Gen-Sel software was used (Fernando and Garrick, 2009), through the bioinformatics platform to implement genomic selection (Bioinformatics to Implement Genomic Selection, BIGS) of the State of Iowa University (USA)¹. The BayesCp option was selected for its advantages over other Bayesian options, when adjusting a model similar to BayesB (Meuwissen *et al.*, 2001; Habier *et al.*, 2011), in addition to estimating the proportion of non-informative *loci* in the model (π), and the assumption of a common variance for all the *loci* with non-zero effects, instead of the individual variance for each *locus* (Habier *et al.*, 2011). The chain used had a length of 41,000, the last 40,000 samples were used to obtain the estimators *a posteriori* of the medium effects and variances, the first 1,000 iterations were burned, and p was fixed at 0.95.

The genomic value of the *i*-th animal was obtained as the sum over all the markers of the *a posteriori* means of the effects predicted from the SNPs, multiplied by the number of copies of allele 1, using the equation:

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

where VGn_i is the genetic molecular value, genomic value, for the *i*-th animal, z_{ij} is the genotype of the *j*-th marker in the *i*-th animal, and $\hat{u}_j$ is the *a posteriori* means of the SNP effect for the *j*-th marker.

The models used to obtain the genomic values in the Gen-Sel were: 1) one without considering the effects of sex and number of generation, and 2) another one that did include them. The criteria for comparison used were accuracy (R^2) and

¹ <http://big.ansci.iastate.edu/>

Cuadro 2. Media y desviación estándar para exactitud (R^2) y varianza del error de predicción (PEV) obtenidas con el modelo completo.
Table 2. Mean and standard deviation for accuracy (R^2) and prediction error variance (PEV) obtained with the complete model.

Generación de evaluación	10	9 y 10	8 a 10	7 a 10
R^2				
11	0.14±0.07	0.28±0.05	0.33±0.05	0.39±0.05
12	0.06±0.08	0.17±0.07	0.23±0.06	0.29±0.07
13	0.01±0.08	0.13±0.07	0.17±0.06	0.23±0.06
14	-0.02±0.09	0.08±0.08	0.13±0.08	0.19±0.07
PEV				
11	0.15±0.01	0.20±0.01	0.22±0.02	0.22±0.02
12	0.17±0.01	0.24±0.02	0.25±0.02	0.26±0.02
13	0.17±0.01	0.25±0.02	0.27±0.02	0.28±0.02
14	0.18±0.01	0.27±0.02	0.29±0.02	0.29±0.02

Los resultados de este estudio son similares a los encontrados por Habier *et al.* (2010), Saatchi *et al.* (2010) y Saatchi *et al.* (2011), quienes observaron que con relaciones cercanas entre los individuos de PEn y PEv se obtienen mejores predicciones, en comparación con los obtenidos cuando estas poblaciones son más distantes.

En congruencia con los resultados para R^2 , el estimador de PEV fue menor, 0.15 ± 0.01 cuando la PEn y la PEv estuvieron más cercanas, independientemente del tamaño de la PEn, generación 10 para PEn y generación 11 para PEv. Al contrario, el mayor PEV (0.29 ± 0.02) se obtuvo cuando la PEn incluyó los individuos de la generación siete y se evaluó la generación 14; es decir cuando era mayor la distancia entre PEn y PEv. En el Cuadro 3 se muestran los estimadores de R^2 y PEV para las diferentes poblaciones de entrenamiento y de evaluación, obtenidos con el modelo que no incluyó los efectos de sexo del animal y generación. Cuando la PEn fue de mayor tamaño (más generaciones incluidas), los estimadores de R^2 fueron mayores. De manera similar, cuando PEv estuvo más cerca de PEn, los estimadores de R^2 fueron mayores.

El estimador de PEV fue menor (0.15 ± 0.01), cuando era menor la distancia entre PEn y PEv, independientemente del tamaño de la PEn, generación 10 para PEn y generación 11 para PEv. Por el contrario, el mayor PEV (0.29 ± 0.02) se obtuvo cuando la PEn incluyó los animales de la generación siete y se evaluó la generación 14. Estos resultados son similares a los

the prediction error variance of the genomic values (PEV). The genetic correlation between the true genetic values simulated obtained with the QMSim software (Sargolzaei and Schenkel, 2009) and the genomic values predicted with the GenSel software (Fernando and Garrick, 2009) were estimated with the ASReml program (Gilmour *et al.*, 2009).

RESULTS AND DISCUSSION

The R^2 estimates were higher when the TP was greater (and included more generations) than the EP. Similarly, the R^2 estimates were higher when the distance between the EP and TP decreased, and they reduced when the distance between both populations increased (Table 2).

The results from this study are similar to those found by Habier *et al.* (2010), Saatchi *et al.* (2010) and Saatchi *et al.* (2011), who observed that with close relationships between individuals from the TP and EP better predictions were obtained, in comparison to those obtained when these populations are more distant.

In agreement with the results for R^2 , the estimator for PEV was smaller, 0.15 ± 0.01 , when the TP and the EP were closer, independently of the size of the TP, generation 10 for TP and generation 11 for EP. On the contrary, the higher PEV (0.29 ± 0.02) was obtained when the TP included the individuals of generation seven and generation 14 was evaluated; that is, when the distance between TP and EP was greatest. Table 3 shows the estimates for R^2 and PEV,

Cuadro 3. Media y desviación estándar para exactitud (R^2) y varianza del error de predicción (PEV), obtenidas con el modelo sin considerar los efectos de sexo del animal y generación.

Table 3. Mean and standard deviation for accuracy (R^2) and prediction error variance (PEV) obtained with the model without taking into consideration the effects of sex and generation of the animal.

Generación de evaluación	10	9 y 10	8 a 10	7 a 10
R^2				
11	0.14±0.07	0.28±0.06	0.34±0.05	0.40±0.04
12	0.06±0.08	0.17±0.06	0.23±0.06	0.30±0.06
13	0.01±0.08	0.12±0.07	0.18±0.07	0.25±0.06
14	-0.03±0.10	0.08±0.08	0.14±0.07	0.19±0.07
PEV				
11	0.15±0.01	0.20±0.02	0.21±0.02	0.22±0.01
12	0.16±0.01	0.24±0.02	0.25±0.02	0.25±0.02
13	0.17±0.01	0.26±0.02	0.27±0.02	0.27±0.02
14	0.18±0.02	0.27±0.02	0.28±0.02	0.29±0.02

encontrados por Habier *et al.* (2010), Saatchi *et al.* (2010) y Saatchi *et al.* (2011), quienes muestran que relaciones cercanas entre los individuos en PEn y PEv tienen mejores predicciones que con grupos más distantes. Los resultados también concuerdan con los de Meuwissen *et al.* (2001) y Hayes *et al.* (2009), quienes señalan que el tamaño de la PEn también influye en la exactitud de la predicción de los VG.

Saatchi *et al.* (2010) obtuvieron aumentos en R^2 de 39.7 y 24.2 % cuando incluyeron las generaciones 51 a 58 y 55 a 58 en la PEn, y la PEv estuvo compuesta por las hembras de la generación 59 y h^2 fue 0.5. Cuando estos autores usaron $h^2=0.1$, los incrementos en R^2 fueron 88.9 % (de 0.112 a 0.2116) y 41.1 % (de 0.112 a 0.2116). Este efecto también fue reportado por Meuwissen *et al.* (2001) y Saatchi *et al.* (2012). Estos resultados pueden ser una expresión de varios factores, como lazos genéticos débiles entre PEn y PEv, un incremento en la tasa de recombinación, un cambio en la estructura genotípica y una reducción en el desequilibrio de ligamiento entre marcadores y QTLs debido a un aumento en el número de generaciones entre ambas poblaciones (Saatchi *et al.*, 2010).

Los estimadores de los coeficientes de correlación genética ($r_{\text{Ghat,GV}}$) entre los valores genómicos predichos (Ghat) con el programa Gen-Sel y los valores genéticos verdaderos (VG), obtenidos con el modelo sin considerar los efectos de sexo y número de

for the different training and evaluation populations, obtained with the model that did not include the effects from animal sex and generation. When the TP was of larger size (more generations included), the R^2 estimators were higher. Similarly, when EP was closer to TP, the R^2 estimators were higher.

The PEV estimate was lower (0.15 ± 0.01), when the distance between TP and EP was lower, regardless of the size of the TP, generation 10 for TP and generation 11 for EP. On the contrary, the higher PEV (0.29 ± 0.02) was obtained when the TP included animals from generation 7 and generation 14 was evaluated. These results are similar to those found by Habier *et al.* (2010), Saatchi *et al.* (2010) and Saatchi *et al.* (2011), who show that close relationships between individuals in TP and EP have better predictions than with more distant groups. The results also agree with those from Meuwissen *et al.* (2001) and Hayes *et al.* (2009), who point out that the size of the TP also influences the accuracy of the prediction of the TGV.

Saatchi *et al.* (2010) obtained increases in R^2 of 39.7 and 24.2 % when they included generations 51 to 58 and 55 to 58 in the TP, and the EP was made up of females from generation 59 and h^2 was 0.5. When these authors used $h^2=0.1$, the increases in R^2 were 88.9 % (from 0.112 to 0.2116) and 41.1 % (from 0.112 to 0.2116). This effect was also reported by Meuwissen *et al.* (2001) and Saatchi *et al.* (2012).

generación y con el modelo completo, son relativamente bajos, aunque es notoria la manera en que su valor disminuye a medida que PEn incluye generaciones más alejadas, de 0.40 a 0.30, y de 0.39 a 0.29 para las PEn que incluyen solamente la generación 10 y para la que incluye de la generación siete a la 10, con los modelos sin considerar los efectos de sexo y número de generación y el completo (Cuadro 4 y 5). Estas correlaciones son menores a las encontradas por Meuwissen *et al.* (2001), quienes reportan valores de 0.318 ± 0.018 a 0.848 ± 0.012 , obtenidos con varios métodos de predicción de valores genómicos (mínimos cuadrados, BLUP, BayesA y BayesB).

Los valores en este estudio son similares a los obtenidos en ganado Hereford por Saatchi *et al.* (2013), con estimadores de correlaciones genéticas entre valores genómicos predichos y DEVG para 10 características de 0.20 ± 0.06 a 0.52 ± 0.04 con el método BayesB, y de 0.18 ± 0.06 a 0.45 ± 0.04 con el método BayesC. Además, estos autores reportan correlaciones genéticas de 0.13 ± 0.06 a 0.32 ± 0.05 entre los valores genéticos deregados y los valores genéticos obtenidos con las metodologías tradicionales de evaluación mediante BLUP. En un estudio de simulación, Spangler *et al.* (2007) obtuvieron para marmoleo de la carne estimadores de correlaciones de 0.256 a 0.859, entre los valores genéticos verdaderos y los predichos, a partir de conjuntos de datos con diferente proporción de información disponible.

Cuadro 4. Estimadores de los coeficientes de correlación genética ($r_{\text{Ghat,GV}}$) entre los valores genómicos predichos con el programa Gen-Sel (Ghat) y los valores genéticos verdaderos (GV), obtenidos con el modelo sin considerar los efectos de sexo y número de generación.

Table 4. Estimates of the genetic correlation coefficients ($r_{\text{Ghat,GV}}$) between the predicted genomic values with the Gen-Sel software (Ghat) and the true genetic values (TGV), obtained with the model without taking into consideration the effects of sex and number of generation.

Población de entrenamiento	Variable		
	Ghat	GV	rGhat,GV
10	0.08	0.36	0.40
9 y 10	0.06	0.37	0.39
8 a 10	0.05	0.37	0.36
7 a 10	0.01	0.40	0.30

Cuadro 5. Estimadores de los coeficientes de correlación genética ($r_{\text{Ghat,GV}}$) entre los valores genómicos predichos (Ghat) con el programa Gen-Sel (Ghat) y los valores genéticos verdaderos (GV), obtenidos con el modelo completo

Table 5. Estimates of the genetic correlation coefficients ($r_{\text{Ghat,GV}}$) between the predicted genomic values (Ghat) with the Gen-Sel software (Ghat) and the true genetic values (TGV), obtained with the complete model.

Población de entrenamiento	Variable		
	Ghat	GV	rGhat,GV
10	0.08	0.38	0.39
9 y 10	0.06	0.38	0.39
8 a 10	0.05	0.38	0.35
7 a 10	0.01	0.39	0.29

These results can be an expression of several factors, such as weak genetic bonds between TP and EP, an increase in the recombination rate, a change in the genotypic structure, and a reduction in the linkage disequilibrium between markers and QTLs due to an increase in the number of generations between both populations (Saatchi *et al.*, 2010).

The estimates for genetic correlation coefficients ($r_{\text{Ghat,GV}}$) are shown, between the genomic values predicted (Ghat) with the Gen-Sel program and the true genetic values (TGV), obtained with the model without taking into consideration the effects of sex and number of generation, and with the complete model, are relatively low, although the way in which their value decreases as the TP includes more distant generations is notable, from 0.40 to 0.30, and from 0.39 to 0.29 for the TP that includes only generation 10 and for the one that includes generation seven to 10, with the models without considering the effects of the sex and number of generation and the complete one, (Table 4 and 5). These correlations are lower than the ones found by Meuwissen *et al.* (2001), who report values of 0.318 ± 0.018 to 0.848 ± 0.012 , obtained with several prediction methods of genomic values (least squares, BLUP, BayesA and BayesB).

The values in this study are similar to those obtained in Hereford cattle by Saatchi *et al.* (2013), with genetic correlation estimates between predicted genomic values and DEVGs for 10 traits of 0.20 ± 0.06 to 0.52 ± 0.04 with the BayesB method, and 0.18 ± 0.06 to 0.45 ± 0.04 with the BayesC

Según Kizilkaya *et al.* (2010), para una característica con heredabilidad de 0.5 y usando aproximadamente 1000 animales en PEn, se obtienen correlaciones mayores de 0.8 para características controladas por aproximadamente 500 *loci*, suponiendo que los *loci* causales son conocidos y genotipados.

La tendencia a disminuir de los estimadores de correlación genética entre los valores genómicos predichos y los valores genéticos verdaderos es clara, en la medida que aumenta la distancia entre PEn y PEV, lo cual sucedió en ambos modelos (Cuadros 4 y 5).

CONCLUSIONES

Las exactitudes obtenidas con los dos modelos de predicción de valores genómicos comparados fueron similares.

Con los dos modelos de predicción comparados, las exactitudes de los valores genómicos predichos fueron mayores en la medida que las poblaciones de entrenamiento incluyeron más generaciones; esto es, que fueron de mayor tamaño. El mismo resultado se obtuvo cuando la distancia entre las poblaciones de entrenamiento y evaluación era menor.

Las varianzas del error de predicción fueron mayores cuando aumentaba la distancia entre las poblaciones de entrenamiento y evaluación.

Los estimadores de las correlaciones genéticas entre los valores genómicos predichos con los modelos comparados y los valores genéticos verdaderos fueron bajos. La disminución de estos estimadores fue notoria a medida que aumentaba la distancia entre las poblaciones de entrenamiento y de evaluación.

LITERATURA CITADA

- Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128:422-428.
- Fernando, R., and D. J. Garrick. 2009. GenSel – user manual for a portfolio of genomic selection related analyses, create 9.1. http://biggs.ansci.iastate.edu/bigsgui/ReleaseNotes/Version-9_1-ReleaseNotes.1.pdf. (Consulta: Enero, 2013).
- Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. *Genet. Sel. Evol.* 43:17-27.
- Garrick, D. J., J. F. Taylor, and R. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55-62.

method. In addition, these authors report genetic correlations of 0.13 ± 0.06 to 0.32 ± 0.05 between the deregressed genetic values and the genetic values obtained with the traditional methodologies of evaluation through BLUP. In a simulation study, Spangler *et al.* (2007) obtained for meat marbling correlation estimates of 0.256 to 0.859, between the true genetic values and the predicted ones, from data sets with different proportion of information available. According to Kizilkaya *et al.* (2010), for one heritability trait of 0.5 and using approximately 1000 animals in the TP, correlations higher than 0.8 are obtained for traits controlled for approximately 500 *loci*, assuming that the causal *loci* are known and genotyped.

The tendency to decrease the genetic correlation estimates between the predicted genetic values and the true genetic values is clear, insofar as the distance between TP and EP increases, which happened in both models (Tables 4 and 5).

CONCLUSIONS

The accuracies obtained with the two prediction models for genomic values compared were similar.

With the two prediction models compared, the accuracy of the predicted genomic values were greater as long as the training populations included more generations; that is, if they were of larger size. The same result was obtained when the distance between the training and evaluation populations was smaller.

The prediction error variances were higher when the distance between the training and evaluation populations increased.

The estimates of genetic correlation between the genomic values predicted with the compared models, and the true genetic values, were low. The decrease in these estimates was notorious insofar as the distance between the training and evaluation populations increased.

—End of the English version—

—*—

- Garrick, D. J., and M. Saatchi. 2011. Opportunities and challenges for genomic selection of beef cattle. *Rev. Bras. Zootec.* 40:310-316.
- Gilmour, A. R., B. J. Gogel, B. R., Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. <http://www.vsnl.co.uk>. (Consulta: Enero, 2013).
- Goddard, M. E., and B. J. Hayes. 1992. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381-391.
- Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323-330.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389-2397.
- Habier, D., J. Tetens, F-R Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 52:5-16.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186-197.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci* 92:433-443.
- Kachman, S. D., M. L. Spangler, G. L. Bennett, K. J. Hanford, L. A. Kuehn, W. M. Snelling, R. M. Thallman, M. Saatchi, D. J. Garrick, R. D. Schnabel, J. F. Taylor, and E. J. Pollak. 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genet. Sel. Evol.* 45:30-38.
- Kizilkaya, K., R. L. Fernando, and D. J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci.* 88:544-551.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157:1819-1829.
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648-4655.
- Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet. Sel. Evol.* 43:38-43.
- Saatchi, M., S. R. Miraei-Ashtiani, A. Nejati Javaremi, M. Moradi-Shahrehabak, and H. Mehrabani-Yeghaneh. 2010. The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *Afr. J. Biotechnol.* 9:438-442.
- Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decjer, T. M. Taxis, R. H. Chapple, H. R. Ramey, S. L. Northcut, S. Bauck, B. Woodward, J. C. M., Dekkers, R. L. Fernando, R. D. Schnabel, D. J. Garrick, and J. F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43:40-55.
- Saatchi, M., R. D. Schnabel, M. M. Rolf, J. F. Taylor, and D. J. Garrick. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genet. Sel. Evol.* 44:38-47.
- Saatchi, M., J. Ward, and D. J. Garrick. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *J. Anim. Sci.* 91:1538-1551.
- Sargolzaei, M., and F. S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25:680-681.
- Spangler, M. L., J. K. Bertrand, and R. Rekaya. 2007. Combining genetic test information and correlated phenotypic records for breeding value estimation. *J. Anim. Sci.* 85:641-649.
- Van Raden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16-24.
- Weller, J. I., M. Shlezinger, and M. Ron. 2005. Correcting for bias in estimation of quantitative trait loci effects. *Genet. Sel. Evol.* 37:501-522.