

Evaluación genómica en ganado Holstein Colombiano, usando genotipos imputados a densidad media

C. Zambrano, Juan; Echeverri, Julián; López Herrera, Albeiro

Evaluación genómica en ganado Holstein Colombiano, usando genotipos imputados a densidad media

Revista MVZ Córdoba, vol. 24, núm. 2, 2019

Universidad de Córdoba, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=69360025024>

DOI: <https://doi.org/10.21897/rmvz.1704>

Evaluación genómica en ganado Holstein Colombiano, usando genotipos imputados a densidad media

Genomic Evaluation of Colombian Holstein Cattle Using Imputed Genotypes at Medium Density

Juan C. Zambrano

Institución Universitaria Colegio Mayor de Antioquia,
Colombia

jczambrano@unal.edu.co

 <http://orcid.org/0000-0001-5205-7621>

DOI: <https://doi.org/10.21897/rmvz.1704>

Redalyc: <http://www.redalyc.org/articulo.oa?id=69360025024>

Julián Echeverri

Universidad Nacional de Colombia, Colombia
jjecheve@unal.edu.co

 <http://orcid.org/0000-0002-9613-0621>

Albeiro López Herrera

Universidad Nacional de Colombia, Colombia
alherrera@unal.edu.co

 <http://orcid.org/0000-0003-1444-3470>

Recepción: 08 Enero 2018

Aprobación: 11 Marzo 2019

Publicación: 10 Junio 2019

RESUMEN:

Objetivo. El objetivo de este estudio fue determinar la precisión y el sesgo de predicción de valores genómicos directos (VGD) usando genotipos imputados a densidad media, en características productivas y reproductivas en ganado Holstein de Antioquia, Colombia. **Materiales y métodos.** Fueron genotipificados 31 animales con el chip Illumina BovineLD, 64 con el chip Illumina BovineSNP50v2 y 48 con el chip Illumina BovineHD. La imputación se realizó usando dos paneles de SNPs (6K y 40K) a una densidad 44K, usando el programa FINDHAP.f90 v4. Los efectos de los SNPs fueron estimados mediante el método bayes C, usando genotipos de baja densidad (6K) y genotipos imputados a una densidad media (44_imputado). La precisión y el sesgo de los VGDs fueron determinados mediante validación cruzada. Las características evaluadas fueron: producción de leche (PL), porcentaje de proteína (PRO), porcentaje de grasa (GRA), puntaje de células somáticas (SCS), intervalo entre partos (IEP) y días abiertos (DA). **Resultados.** Las precisiones de VGD (rpVGD;EBV) en todas las características evaluadas oscilaron entre 0.19 y 0.24 y el sesgo (bVGD;EBV) entre 0.03 y 0.16 cuando se usó el panel 6K y usando el panel 44K_imputado las precisiones fueron mayores, oscilado entre 0.24 y 0.33 y sesgo entre 0.03 y 0.26. **Conclusiones.** La precisión de predicción de los VGDs fue mayor cuando se usaron genotipos imputados a densidad media, en comparación con la precisión de predicción obtenida empleando genotipos de baja densidad. Por lo cual, en este estudio se concluye que la imputación de genotipos es muy útil dado que aumenta la confiabilidad de la evaluación genómica.

PALABRAS CLAVE: Genotipificación, polimorfismo de nucleótido simple, selección genómica.

ABSTRACT:

Objective. The goal of this study was to determine the accuracy and bias of direct genomic values (DGV) using imputed genotypes at medium density in yield- and reproduction-related traits for Holstein cattle from Antioquia, Colombia. **Materials and Methods.** A total of 31 animals were genotyped with the Illumina BovineLD chip, 64 with Illumina BovineSNP50v2 and 48 with Illumina BovineHD. Two SNP panels (6K and 40K) were imputed to a density of 44K using the FINDHAP.f90 v4 program. The effects of the SNPs were estimated using the Bayes C method, using low-density (6K) genotypes as well as medium-density imputed genotypes (44_imputed). The accuracy and bias of the DGVs were determined by cross-validation. The evaluated traits were: milk yield (MY), percentage of protein (PP), percentage of fat (PF), somatic cell score (SCS), calving interval (CI) and open days (OD). **Results.** When using the 6K panel, the accuracy values for DGV (rpDGV;EBV) in all the studied traits ranged from

0.19 to 0.24, and the bias (bDGV;EBV) from 0.03 to 0.16. In contrast, using the 44K_imputed panel generated higher accuracy values ranging from 0.24 to 0.33 and a bias ranging from 0.03 to 0.26. **Conclusions.** The accuracy of prediction the DGV was higher with genotypes imputed to medium densities when compared to the accuracy of prediction obtained using low-density genotypes. Therefore, in this study it is concluded that the imputation of genotypes is very useful, because it improves the reliability of the genomic evaluation.

KEYWORDS: Genotyping, genomic selection, single nucleotide polymorphism .

INTRODUCCIÓN

La selección genómica (SG) es una metodología que incluye miles de marcadores moleculares, los cuales cubren todo el genoma uniformemente; de tal manera que cada QTL asociado a una característica en particular, esté en desequilibrio de ligamiento con al menos un marcador (1,2). Esta metodología combina la información de fenotipos (o pseudofenotipos: EBVs), el pedigree y el genotipo de cada animal, para estimar valores de cría genómicos (GEBVs), los cuales son el punto central en los actuales esquemas de mejoramiento genético (3). Esto ha generado nuevas oportunidades para realizar una selección más eficiente en características que incluso presentan baja heredabilidad o que son difíciles de medir (4,5).

Sin embargo, el mayor obstáculo en la implementación de la SG, es el costo de genotipificación de animales, principalmente con chips de moderada y alta densidad (4). Para reducir los costos de genotipificación, una estrategia es usar chips de SNPs de baja densidad, los cuales pueden ser optimizados por procesos de imputación. Las estrategias de genotipificación para imputación generalmente involucran animales influyentes, principalmente machos genotipificados con chips de alta densidad (Illumina SNP50, Illumina BovineHD 800K), animales jóvenes (candidatos a selección) y vacas, usando chips de baja densidad (Illumina BovineLD) y en algunos casos individuos no genotipificados (6,7,8). La imputación de genotipos inicia con la reconstrucción de haplotipos portados por los individuos genotipificados con chips de alta densidad. Luego los genotipos de baja densidad son usados en conjunto con el pedigree, el ligamiento familiar y la información de desequilibrio de ligamiento para determinar la combinación de haplotipos presentes en los individuos genotipificados con chips de baja densidad y deducir su genotipo a una densidad de SNPs mayor que la inicial (8).

Varios programas de imputación han sido creados como: fastPHASE (9), BEAGLE (10), IMPUTE2 (11), FINDHAP (12), FIImpute (13), entre otros. Todos los programas usan diferentes métodos de reconstrucción de haplotipos y por lo tanto el tiempo de computación es diferente. FastPHASE, BEAGLE e IMPUTE2, son lentos ya que usan métodos bayesianos, los cuales pueden ser limitantes en la práctica cuando se usan grandes cantidades de datos. FINDHAP y FIImpute usan métodos determinísticos y computacionalmente son muy rápidos y además las precisiones de imputación son comparables con fastPHASE, BEAGLE e IMPUTE2 (6,14). La imputación de genotipos en selección genómica, ha permitido estimar GEBVs en toros y vacas, con confiabilidades similares a las obtenidas, si los animales fueran genotipificados con chips de alta densidad (6,14). Esto ha facilitado la implementación de la SG en programas de mejoramiento genético, haciendo que esta metodología sea más confiable, más eficiente y menos costosa (5,15,16).

A pesar de que en Colombia existe una alta tecnificación en ganado lechero especializado, el mejoramiento genético ha presentado serias dificultades, debido a la limitada cantidad de registros productivos disponibles (17). Sin embargo, esto no ha sido un impedimento para iniciar con la implementación de metodologías que incluyen información genómica tanto en ganado de leche (18,19), en razas criollas (20), así como en ganado de carne (21). El objetivo de esta investigación fue determinar la precisión y el sesgo de valores genómicos directos (VGD) usando genotipos imputados a densidad media, en características productivas y reproductivas en ganado Holstein de Antioquia, Colombia.

MATERIALES Y MÉTODOS

Población. Esta investigación fue realizada con base en la información obtenida de 85 hatos lecheros para la raza Holstein, ubicados en 18 municipios del Departamento de Antioquia (Colombia) que cubre zonas de bosque muy húmedo montano bajo (bmh-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 de manejo, alimentación y sanidad fueron variables en todos los hatos, así como su topografía y ubicación geográfica.

El número de registros utilizados para los análisis fue variable, de acuerdo con la característica dependiente analizada: 8772, 6624, 6595, 6426, 11562 y 11395 para producción de leche (PL), porcentaje de proteína (PRO), porcentaje de grasa (GRA), recuento de células somáticas (RCS), intervalo entre partos (IEP) y días abiertos (DA), 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 (22). El número de animales en el pedigree para la evaluación genética convencional fue de 9090 (que incluyó 721 padres y 944 madres) y para evaluación genómica fue de 144 (36 toros y 108 vacas).

Genotipificación de animales con chips de alta densidad. Fueron genotipificados 144 animales con tres chips de Illumina (Illumina Inc, San Diego, CA): 31 con el chip Bovine LD, 65 con el chip BovineSNP50 v2 y 48 con el chip BovineHD. Después de aplicar el control de calidad: call rate > 0.90, alelos de menor frecuencia (MAF) > 0.02, equilibrio de Hardy Weinberg (HWE) $p > 0.001$ y descarte de marcadores con errores mendelianos, fueron obtenidos un total de 6716 marcadores comunes a los tres chips, cuya densidad fue denominada 6K. La edición de los genotipos fue realizada con los programas SAS versión 9.2 (SAS Institute Inc., Cary, E.E.U.U.) y PLINK v1.07 (CHGR., Massachusetts, E.E.U.U.)

Imputación de genotipos. Para obtener la base de datos de genotipos imputados, fueron seleccionados 6820 SNPs (panel 7K) del chip LD y 40321 SNPs (panel 40K) del chip HD, los cuales fueron comunes al chip SNP50v2. Los paneles 7K y 40K fueron imputados a 44224 SNPs (panel 44K_imputado). Los SNPs fueron definidos como genotipos discretos (0, 1 y 2), por lo cual se recodificaron como: BB=0 (homocigótico para el primer alelo), AB=1 (heterocigótico), AA=2 (homocigótico para el segundo alelo), B_=3 (alelo paterno conocido y alelo materno no conocido), _A=4 (Alelo paterno no conocido y alelo materno conocido) y 5 cuando ambos alelos son no conocidos. Los SNPs fueron ordenados por número de cromosoma y localización sobre el cromosoma, teniendo en cuenta el ensamble UMD3.1 del genoma bovino. El pedigree de los 144 animales genotipificados que correspondió a 761 animales, fue incluido en el análisis, con el fin de mejorar la exactitud de la imputación.

La exactitud de la imputación fue determinada por comparación de los genotipos imputados (GI) con los genotipos verdaderos (GV) (genotipos obtenidos con el chip SNP50v2), mediante la correlación de Pearson ($r_{GV;GI}$). El proceso de imputación fue llevado a cabo mediante el programa FINDHAP.f90 versión 4 (USDA., Beltsville, E.E.U.U.).

Estimación de los efectos de los marcadores y valores genómicos directos. La estimación de los efectos de los SNPs fue realizada empleando el método Bayes C (23). El modelo estadístico general fue descrito como:

$$y = 1_n \mu + \sum_{i=1}^p X_{ij} g_j \delta_j + Zu + e \quad [\text{Ecuación.}]$$

Donde: y es el vector de valores fenotípicos corregidos por efectos fijos (que corresponde a los EBVs convencionales obtenidos mediante el método BLUP), para las características PL, SCS, PRO, GRA, DA e IEP; μ corresponde a la media general, 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 (BB), el heterocigótico (AB) y el homocigótico del otro alelo (AA). 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 $\sigma^2 g$ y se asume normalmente distribuido $N \sim (0, \sigma^2 g)$ cuando $\delta_j = 1$, pero $g_j = 0$ cuando $\delta_j = 0$. δ_j es una variable aleatoria 0/1 que indica ausencia (con probabilidad π) 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 fue considerada 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, $u \sim N(0, A\sigma^2 u)$, donde A es la matriz de parentesco derivada del pedigree de los animales genotipificados, e es el residual y también se asume estar distribuido normalmente, $e \sim N(0, I\sigma^2 e)$, 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 π) 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 (24).

Así, los efectos de los SNPs (g_j) siguen una distribución *a priori* normal con media cero y varianza ($\sigma^2 g$), mientras que las varianzas de los efectos de los marcadores ($\sigma^2 g$) siguen una distribución *a priori* Chi cuadrada invertida escalada con parámetros n_g (grados de libertad) y s^2_g (parámetro de escala). Por otra parte, π es tratada como una distribución *a priori* uniforme con media cero y varianza 1 (24), como se describe a continuación:

$$g_j \left| \begin{array}{l} \pi, \sigma^2 g \\ g_j = 0 \end{array} \right. \begin{array}{l} \xrightarrow{\sim N(0, \sigma^2 g)} \text{conprobabilidad } 1-\pi \\ \xrightarrow{\quad} \text{conprobabilidad } \pi \end{array} \quad \text{[Ecuación.]}$$

$$\sigma^2 g | V_g S_g^2 \sim V_g S_g^2 X^{-2} \quad \text{[Ecuación.]}$$

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

El VGD de los animales cuyo genotipo fue conocido, se determinó como:

$$y = \sum_{i=1}^p X_{ij} \hat{g}_j \delta_j \quad \text{[Ecuación.]}$$

Corresponde a la suma de todos los efectos de los SNPs estimados (\hat{g}_j). El programa utilizado para la estimación de los efectos de los SNPs y de los VGDs fue GS3 (INRA., Toulouse, Francia).

Validación cruzada. La precisión de la selección genómica fue determinada mediante validación cruzada, como lo describe Meuwissen et al. (25). La base de datos completa correspondiente a 144 animales con genotipo y fenotipo conocidos fue dividida aleatoriamente en $S=12$ grupos de igual tamaño (S_1-S_{12}). En la primera etapa, S_1 correspondió al grupo de validación, en el cual los animales presentaron genotipo pero no fenotipo ya que fue removido de la base de datos (fenotipo enmascarado=0), mientras que de S_2 a S_{12} correspondieron al grupo de referencia (con fenotipo y genotipo), a partir del cual se estimaron los efectos de los SNPs, en S_1 fueron estimados los VGDs de los 12 animales que conforman el grupo, usando los efectos de los SNPs estimados en los 132 animales remanentes (S_2-S_{12}). Esto se repitió para cada uno de los grupos, hasta llegar al grupo S_{12} . Al final se estimaron los VGDs de todos los animales.

Precisión y sesgo de los VGD predichos. Los VGDs estimados mediante el método bayes C fueron comparados con los valores genéticos convencionales (EBV) mediante la correlación de Pearson, que fue considerada como la precisión de la selección genómica (26). La ecuación se representa como:

$$r_p = \frac{COV(x, y)}{\sigma_x \sigma_y} \quad [\text{Ecuación.}]$$

Donde, r_p = coeficiente de correlación de Pearson, x =VGD para la característica i , y =EBV convencional para la característica i . También fue determinado el coeficiente de correlación por rangos de Spearman, el cual fue usado para determinar el grado de similitud entre las clasificaciones o ranking de los animales por sus valores genómicos (VGD o GEBV) y la clasificación de los mismos animales por el EBV convencional.

La ecuación empleada fue descrita como:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \quad [\text{Ecuación.}]$$

Donde:

D = diferencia entre los correspondientes valores genéticos de orden ($x-y$), x = VGD, 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 genéticos (VGD y EBV convencional) es similar.

Finalmente, el coeficiente de regresión lineal de los EBVs convencionales sobre los VGDs ($b_{VGD;EBV}$) fue determinado. El $b_{VGD;EBV}$ se consideró como el sesgo de la predicción de VGD. La regresión lineal fue definida como:

$$y_i = b_0 + b_1 x_i + e \quad [\text{Ecuación.}]$$

Donde y_i = variable dependiente que corresponde EBV convencional para la característica i , x_i = VGD para la característica i , b_1 = coeficiente de regresión del EBV convencional sobre el VGD, b_0 = intercepto y e = residual.

RESULTADOS

Análisis descriptivo para características productivas y reproductivas. Fueron determinados los promedios para las características productivas PL, PRO, GRA y SCS y para las características reproductivas IEP y DA (Tabla 1), que son resultados normales para la raza Holstein en condiciones del trópico alto colombiano. Por otra parte, las heredabilidades estimadas en la evaluación genética convencional para características productivas (GRA, PRO, PL y SCS) fueron moderadas con valores de 0.33, 0.41, 0.30 y 0.27 respectivamente y para las características reproductivas IEP y DA fueron bajas con valores de 0.11 en los dos casos.

TABLA 1
Análisis descriptivo para características productivas y reproductivas en ganado Holstein de Antioquia, Colombia.

Característica	N	Media±DE	h^2 (EE)
PL	8772	5400±2126	0.30(0.017)
PRO	6624	3.10±0.31	0.33(0.020)
GRA	6595	3.97±0.55	0.41(0.019)
SCS	6426	4.62±1.37	0.27(0.021)
IEP	11562	410±78	0.11(0.011)
DA	11395	127±77	0.11(0.011)

PL - producción de leche (en Litros/lactancia), GRA - porcentaje de grasa, PRO - porcentaje de proteína, SCS - puntaje de células somáticas, IEP - intervalo entre partos, DA - días abiertos, N - número de registros, DE - desviación estándar, h^2 - heredabilidad, EE - Error estándar.

Precisión de la imputación. La base de datos 7K conformada por 31 animales fue imputada a 44K, obteniendo una exactitud de 0.73. De igual manera, la base de datos 44K conformada por 48 animales fue imputada a 44K, logrando una exactitud de 0.97 (Tabla 2).

TABLA 2
Precisiones de imputación de genotipos

Población	Chip	Número de animales	Densidad (SNPs)	Densidad Imputada	Precisión de imputación (R^2)
144	BovineLD 31	6820 (7K)	44224	0.730	
	BovineHD 48	40321 (40K)	44224	0.970	
	SNP50K 65	44224 (44K)	44224	0.998	

Ánalisis descriptivo para los efectos de los marcadores. Los efectos de los SNPs fueron estimados en todas las características incluídas en el análisis (PL, PRO, GRA, SCS, IEP y DA), usando genotipos de baja densidad (6K) y genotipos imputados (44K_imputado). Los efectos de los SNPs fueron mayores en todas las características cuando se usó la base de datos 6K, en comparación con los efectos de los SNPs estimados en la base de datos 44K_imputado (Tabla 3).

TABLA 3
Análisis estadístico descriptivo de los efectos de los SNPs para características productivas y reproductivas en ganado Holstein de Antioquia, Colombia.

Característica	Panel	Densidad (SNP)	Media±DE	Mínimo	Máximo
PL			-6.2×10^{-2} ±0.610	-3.75	3.57
PRO			1.0×10^{-5} ±0.0012	-0.009	0.008
GRA			7.5×10^{-6} ±0.0018	-0.016	0.013
SCS	6K	6716	4.0×10^{-5} ±0.0009	-0.004	0.005
IEP			3.0×10^{-4} ±0.027	-0.115	0.105
DA			-6.2×10^{-4} ±0.056	-0.253	0.266
PL			1.4×10^{-3} ±0.507	-1.95	2.54
PRO			8.5×10^{-7} ±0.0003	-0.002	0.002
GRA			-3.8×10^{-8} ±0.0003	-0.002	0.002
SCS	44K_imputado 44224		2.8×10^{-7} ±0.00018	-0.001	0.001
IEP			2.0×10^{-5} ±0.003	-0.017	0.019
DA			7.3×10^{-5} ±0.016	-0.081	0.076

PL - producción de leche (en Litros/lactancia), GRA - porcentaje de grasa, PRO - porcentaje de proteína, SCS - puntaje de células somáticas, IEP - intervalo entre partos, DA - días abiertos, DE - Desviación estándar.

Precisión de los valores genómicos directos. Las precisiones genómicas ($r_{pVGD;EBV}$) oscilaron entre 0.19 y 0.29, cuando se usó la base de datos de genotipos 6K, mientras que cuando se usó la base de datos 44K_imputado, las precisiones genómicas fueron mayores y oscilaron entre 0.24 y 0.33 en todas las características evaluadas. Aunque las correlaciones obtenidas fueron bajas, tanto usando genotipos de baja densidad, como genotipos de alta densidad, puede evidenciarse un incremento de las precisiones cuando se usa una densidad mayor de SNPs y es más notorio en características de mayor heredabilidad (Tabla 4). De igual manera, las correlaciones de Spearman ($r_{sVGD;EBV}$) oscilaron entre 0.16 y 0.33 cuando se usó la base de datos 6K, mientras que cuando se usaron datos imputados (44_imputado), las correlaciones fueron mayores, con valores que oscilaron entre 0.30 y 0.35, excepto en IEP que presentó un valor de 0.20 (Tabla 4).

TABLA 4

Precisión genómica de los valores genómicos directos y correlación de Spearman entre el VGD y el EBV.

Característica	6K	44K_imputado	6K	44K_imputado
	r_p VGD;EBV \pm EE	r_p VGD;EBV \pm EE	r_s VGD;EBV \pm EE	r_s VGD;EBV \pm EE
PL	0.24 \pm 0.081	0.32 \pm 0.080	0.27 \pm 0.081	0.30 \pm 0.080
PRO	0.24 \pm 0.081	0.33 \pm 0.079	0.16 \pm 0.083	0.35 \pm 0.079
GRA	0.29 \pm 0.080	0.33 \pm 0.079	0.33 \pm 0.079	0.38 \pm 0.078
SCS	0.22 \pm 0.082	0.30 \pm 0.080	0.22 \pm 0.082	0.32 \pm 0.080
IEP	0.19 \pm 0.082	0.24 \pm 0.081	0.28 \pm 0.081	0.20 \pm 0.082
DA	0.19 \pm 0.082	0.26 \pm 0.081	0.18 \pm 0.083	0.21 \pm 0.082

PL - producción de leche, GRA - porcentaje de grasa, PRO - porcentaje de proteína, SCS - puntaje de células somáticas, IEP - intervalo entre partos, DA - días abiertos, r_p VGD;EBV - correlación de Pearson entre VGD y EBV, EE - error estándar, r_s VGD;EBV - correlación de Spearman entre VGD y EBV

Los coeficientes de regresión ($b_{VGD;EBV}$) obtenidos para PL, PRO, GRA, SCS, IEP y DA fueron bajos: 0.16, 0.03, 0.11 y 0.11 respectivamente, cuando se usó la base de datos 6K, mientras que cuando se usó la base de datos 44K_imputado, los coeficientes fueron mayores solo en PL, PRO, GRA y SCS con valores de: 0.26, 0.12, 0.20 y 0.16 respectivamente. Para IEP y DA no presentaron cambios significativos, con valores de 0.03 y 0.04 respectivamente (Tabla 5).

TABLA 5

Sesgo de predicción de los EBVs convencionales sobre los VGDs para características productivas y reproductivas en ganado Holstein de Antioquia, Colombia.

Característica	6K	44K_imputado
	b VGD;EBV \pm EE	b VGD;EBV \pm EE
PL	0.16 \pm 0.054	0.26 \pm 0.064
PRO	0.03 \pm 0.011	0.12 \pm 0.030
GRA	0.11 \pm 0.030	0.20 \pm 0.048
SCS	0.11 \pm 0.042	0.16 \pm 0.044
IEP	0.03 \pm 0.013	0.03 \pm 0.010
DA	0.03 \pm 0.015	0.04 \pm 0.013

PL - producción de leche, GRA - porcentaje de grasa, PRO - porcentaje de proteína, SCS - puntaje de células somáticas, IEP - intervalo entre partos, DA - días abiertos, b VGD;EBV - coeficiente de regresión del EBV convencional sobre el VGD (sesgo de predicción).

DISCUSIÓN

En la presente investigación, se determinó la precisión de imputación de genotipos, obteniendo valores de 0.97 cuando se imputó de 40K a 44K, mientras que la precisión fue menor (0.73), cuando se imputó genotipos de 6K a 44K. Khatkar et al (6), determinaron precisiones de imputación usando los programas IMPUTE2, Beagle y Fastphase con valores de 0.8983, 0.8595, y 0.7630, cuando imputaron genotipos de 3K a 50K y 0.9731, 0.9597 y 0.9626 cuando imputaron de 35K a 50K respectivamente para cada programa, usado una población de referencia de 136 animales; cuando la población de referencia fue de 27 animales, las precisiones de imputación fueron menores con valores de 0.8160, 0.6598 y 0.7409 cuando imputaron de 3K a 50K y 0.9229, 0.8543 y 0.9351 al imputar de 35K a 50K. En otra investigación, Chen et al (26), determinaron

las precisiones de imputación para las densidades de SNPs: 6K, 3K, 1536SNP, 768SNP y 384SNP, cuando fueron imputadas a 50K, usando el programa FIMpute2, obteniendo valores de 0.9723, 0.9367, 0.9120, 0.8285 y 0.7210 respectivamente para cada densidad. En esta investigación los autores incluyeron 2046 animales en la población de referencia (genotipos 50K). Estos resultados concuerdan con los resultados de la presente investigación y en todos los casos se sugiere que el número de animales en la población de referencia y la densidad de los marcadores son factores que afectan significativamente la precisión de imputación.

Por otra parte, el uso de paneles de SNPs de alta densidad puede incrementar la precisión de la selección genómica, mucho más que con paneles de baja densidad (6,27). Sin embargo, el costo de genotipificación de animales con chips de alta densidad es mucho más elevado, por lo cual, se hace necesario el uso de datos imputados, para mejorar la predicción de valores genómicos. En la presente investigación, se usaron genotipos de baja densidad (6K) y genotipos imputados (44K_imputado), en los cuales se determinó la precisión (r^2 VGD;EBV) y el sesgo de predicción (b^2 VGD;EBV), para seis características de importancia económica (PL, PRO, GRA, SCS, IEP y DA). Las precisiones de predicción de VGD fueron en promedio 0.23 cuando se usó el panel 6K y 0.30 cuando se usó el panel 44K_imputado.

Khatkar et al (6), igualmente determinaron la precisión de los VGDs en una población de validación de 452 toros, usando los efectos de los SNPs de una población de referencia de 1753 animales. La precisión fue calculada usando los genotipos LD, 50K_imputado y 800K_imputado. Las precisiones para PL ($h^2 = 0.25$) fueron: 0.481, 0.546 y 0.558 y para fertilidad ($h^2 = 0.04$) fueron menores: 0.232, 0.246 y 0.256 respectivamente para cada densidad. Los resultados indican que la precisión fue mayor usando la densidad 800K_imputado, comparada con las precisiones obtenidas para 50K_imputado y LD. Por otra parte, Chen et al (26), determinaron mejores precisiones cuando usaron paneles con mayor densidad de SNPs para PL y PRO. Para PL las precisiones fueron: 0.64 (50K), 0.61 (3K), 0.58 (L1536), 0.52 (L768) y 0.42 (L384). Para PRO las precisiones fueron: 0.76 (50K), 0.72 (3K), 0.59 (L1536), 0.51 (L768) y 0.34 (L384). En la presente investigación, las precisiones obtenidas para PL fueron 0.24 y 0.32, para PRO 0.34 y 0.33, para IEP 0.19 y 0.24 y para DA 0.19 y 0.26 cuando se usaron los paneles 6K y 44K respectivamente. En la investigación de Chen et al (26), determinaron un incremento promedio en la precisión de 7 puntos cuando pasaron de 1536 a 6177 SNPs y de 7.8 puntos cuando pasaron de 1536 a 35790 SNPs, para cuatro características evaluadas (PL, PRO, GRA y SCS). En la presente investigación se determinó un incremento en la precisión de 6.8 puntos en promedio para las mismas características (PL, PRO, GRA y SCS), cuando se pasó del panel 6K al panel 44K_imputado, lo que indican que la densidad de los marcadores, juega un papel crucial en la precisión de la predicción genómica, tanto en características de alta como de baja heredabilidad, lo que muestra la importancia de usar chip de moderada o alta densidad y/o de realizar imputación de genotipos. La explicación a este aumento en la precisión de predicción de VGD se debe a que un panel con mayor número de SNPs permite cubrir una mayor región en el genoma, y por lo tanto, capturar un mayor número de QTLs involucrados en la expresión de las características cuantitativas.

En la actual investigación, las precisiones de predicción de los VGD para PRO y GRA fueron iguales con un valor de 0.33 para las dos características usando el panel 44K. De igual manera, Chen et al (26), determinaron las precisiones de predicción de VGD usando un panel de 50K, para PRO y GRA, obteniendo valores de: 0.75 y 0.76 respectivamente. En las dos investigaciones, se puede ver que PRO y GRA tienen una arquitectura genética similar, es decir que hay pocos QTLs con gran efecto, comparadas con otras características como SCS donde hay muchos QTLs con efecto pequeño. Ahora, la diferencia en las precisiones en las dos investigaciones está asociada al tamaño de la población referencia. En la presente investigación fueron usados 132 animales, mientras que en la investigación de Chen et al (26) fueron usados 10309 animales. Además, las metodologías usadas fueron diferentes, Bayes B en el primer caso y Bayes C en la actual investigación. En Bayes B cada SNP tiene una varianza locus específica, mientras que en Bayes C, la varianza es común para todos los SNPs (23).

Habier et al (28), determinaron la precisión de predicción de GEBVs en 113 toros jóvenes de raza Holstein, usando los efectos de 40764 SNPs estimados en una población de referencia de 4000 animales. Las precisiones

obtenidas ($r_{pGEBV;DEBV}$) para PL y SCS fueron 0.46 y 0.18 usando el método bayes A, 0.41 y 0.12 usando el método bayes B y 0.43 y 0.15 usando el método bayes Cpi. Para características de baja heredabilidad como SCS, muchos QTL con efectos pequeños no pueden ser detectados, ya que el desequilibrio de ligamiento entre marcadores y QTL puede ser muy bajo, por lo cual se requiere una mayor densidad de marcadores. Con un panel 50K, no ha sido posible aún capturar el 100% de la varianza genética en todas las características evaluadas en ganado lechero, el valor capturado está entre el 90% para PL a un 32% para características de fertilidad (28).

En otra investigación, Hayes et al (29), determinaron la confiabilidad (definida mediante la relación $r_{GEBV;EBV}/r_{EBV;TEBV}$) de valores genómicos en animales jóvenes usando el método bayes A, para PRO y fertilidad, usando una población de referencia de 332 toros, a partir de los cuales se estimaron los efectos de 38259 SNPs. La confiabilidad para PRO fue de 0.36, mientras que para fertilidad fue de 0.14. Por su parte, Nicolazzi et al (30) determinaron la precisión de predicción de valores genómicos ($r_{ABV;DGV}$), usando los métodos Bayes A y Bayes LASSO-gamma para producción de proteína (PPRO) y GRA en una población de validación de 386 toros, usando los efectos de 39048 SNPs estimados en una población de referencia de 763 toros de raza Holstein-Friesian. Las precisiones de predicción obtenidas para PPRO fueron 0.52 y 0.48 respectivamente para cada método, mientras que para GRA los valores fueron 0.75 y 0.71 respectivamente. Los autores sugieren que los resultados de precisión de predicción son mejores para GRA, dado que esta característica es explicada en más de un 30% por mutaciones en los genes DGAT1 y GHR, indicando que GRA es una característica más heredable que PPRO. Colombani et al (31), también determinaron la precisión y el sesgo de los valores genómicos en las razas Holstein (2976 toros en la población de referencia) y Montbéliarde (950 toros en la población de referencia), usando el método bayes Cpi. Los SNPs evaluados fueron 38462 para Montbéliarde y 39738 para Holstein. Las precisiones para PL, GRA y tasa de concepción (TC) fueron: 0.57, 0.80 y 0.34 para raza Holstein y 0.44, 0.62 y 0.43 respectivamente para la raza Montbéliarde. Los coeficientes de regresión (sesgo de predicción) fueron: 0.73, 0.90 y 0.72 en Holstein y 0.74, 0.85 y 1.35 en la raza Montbéliarde.

En la actual investigación, las precisiones obtenidas para PRO y GRA fueron mayores (0.33 en los dos casos) comparadas con las obtenidas para DA e IEP que fueron 0.24 y 0.26 respectivamente. De igual manera, el coeficiente de regresión fue mayor para PRO y GRA (0.12 y 0.20) comparado con, IEP y DA que fueron 0.03 y 0.04 respectivamente. Estos coeficientes en todos los casos fueron menores de 1 ($b < 1$) indicando que los VGD fueron sobreestimados con respecto a los EBVs convencionales, lo que sugiere una inflación de la varianza genética en todas las características. Hayes et al (29), hacen referencia a que la confiabilidad de los GEBVs para fertilidad fue baja comparada con PRO, considerando la alta diferencia de heredabilidad entre características evaluadas y sugieren que se requiere una mayor cantidad de registros para características reproductivas en experimentos iniciales para obtener GEBV con mayor confiabilidad y lograr que la selección genómica sea más eficiente.

Finalmente, debe considerarse que en la presente investigación fueron estimados los efectos de los marcadores en una población de referencia pequeña, por lo cual las precisiones y los coeficientes de regresión fueron bajos comparados con los resultados de las investigaciones de los autores mencionados anteriormente. Por lo tanto, es importante aclarar que esta investigación muestra resultados iniciales de evaluaciones genómicas para características de interés económico tanto de baja como de alta heredabilidad en el trópico alto colombiano. No obstante, es necesario incrementar el número de animales genotipificados y el número de SNPs para lograr consolidar una población referencia de la cual se pueda realizar evaluaciones genómicas con mayor precisión.

Comité de ética

El protocolo experimental de esta investigación fue aprobado por el Comité de Ética en Investigación de la Universidad Nacional de Colombia sede Medellín (Número de la carta de aprobación: CEMED-015 Mayo, 2012).

Conflictos de interés

Ninguno de los autores tiene conflicto de intereses en relación con a este artículo

Agradecimientos

Sinceros agradecimientos a la Universidad Nacional de Colombia sede Medellín, a la Cooperativa COLANTA LTDA y al Ministerio de Agricultura y Desarrollo Rural del gobierno del Colombia por su financiación de esta investigación.

REFERENCIAS

1. Goddard ME, Hayes BJ. Genomic Selection. *J Anim Breed Genet.* 2007; 124(6):323-330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
2. Wang L, Zhu G, Johnson W, Kher M. Three new approaches to genomic selection. *Plant Breeding.* 2018;137(5):673–681. <https://doi.org/10.1111/pbr.12640>
3. Meuwissen TH, Hayes BJ, Goddard ME: Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001; 157(4):1819–1829. <https://www.ncbi.nlm.nih.gov/pubmed/11290733>
4. Boichard D, Chung H, Dassonneville R, David X, Eggen A, Fritz S, et al. Design of a bovine low-density SNP array optimized for imputation. *PLoS ONE.* 2012; 7(3):e34130. <https://doi.org/10.1371/journal.pone.0034130>
5. Weng Z, Zhang Z, Ding X, Fu W, Ma P, Wang C, Zhang Q. Application of imputation methods to genomic selection in Chinese Holstein cattle. *J Anim Sci Biotechnol.* 2012; 3(1):6. <https://doi.org/10.1186/2049-1891-3-6>
6. Khatkar MS, Moser G, Hayes BJ, Raadsma HW. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics.* 2012; 13(1):538. <https://doi.org/10.1186/1471-2164-13-538>
7. Schefers J, Weigel KA. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Anim Front.* 2012; 12(1):4-9. <https://doi.org/10.2527/af.2011-0032>
8. Huang YJ, Hickey JM, Cleveland MA, Maltecca C. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet Sel Evol.* 2012; 44(1):25. <https://doi.org/10.1186/1297-9686-44-25>
9. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 2006; 78(4):629-644. <https://doi.org/10.1086/502802>
10. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 2009; 84(2):210-223. <https://doi.org/10.1016/j.ajhg.2009.01.005>
11. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5(6):e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
12. VanRaden PM, Null DJ, Sargolzaei M, Wiggans GR, Tooker ME, Cole JB, et al. Genomic imputation and evaluation using high-density Holstein genotypes. *J Dairy Sci.* 2013; 96(1):668–678. <https://doi.org/10.3168/jds.2012-5702>
13. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 2014; 15: 478. <https://doi.org/10.1186/1471-2164-15-478>
14. Weigel KA, Van Tassell CP, O'Connell JR, VanRaden PM, Wiggans GR. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. *J Dairy Sci.* 2010; 93(5):2229-2238. <https://doi.org/10.3168/jds.2009-2849>
15. Zhang Z, Druet T. Marker imputation with low-density marker panels in Dutch Holstein cattle. *J Dairy Sci.* 2010; 93(11):5487-5494. <https://doi.org/10.3168/jds.2010-3501>

16. Wiggans GR, Cole JB, Hubbard SM, Sonstegard TS. Genomic Selection in Dairy Cattle: The USDA Experience. *Ann Rev Anim Biosci.* 2017; 5(1):309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>
17. Rincón JC, Zambrano JC, Echeverri JJ. Estimation of genetic and phenotypic parameters for production traits in Holstein and Jersey from Colombia. *Rev MVZ Córdoba.* 2015; 20(Supl):4962-4973. <https://doi.org/10.21897/rmvz.11>
18. Echeverri J, Zambrano JC, López-Herrera A. Genomic evaluation of Holstein Cattle in Antioquia (Colombia): a case study. *Rev Colomb Cienc Pecu.* 2014; 27(4):306-314. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-06902014000400009
19. Zambrano JC, Rincón JC, López A, Echeverri JJ. Estimation and comparison of conventional and genomic breeding values in Holstein cattle of Antioquia, Colombia. *Rev MVZ Córdoba.* 2015; 20(3):4739-4753. <https://doi.org/10.21897/rmvz.44>
20. Martínez R, Gómez Y, Rocha JFM. Genome-wide association study on growth traits in Colombian creole breeds and crossbreeds with Zebu cattle. *Genet Mol Res.* 2014; 13(3):6420-6432. <https://doi.org/10.4238/2014.august.25.5>
21. Martínez R, Mar JF, Bejarano D, Burgos W. Genomic predictions and accuracy of weight traits in a breeding program for Colombian Zebu Brahman [On line]. Proceedings of the World Congress on Genetics Applied to Livestock Production. 2018. <http://www.wcgalp.org/system/files/proceedings/2018/genomic-predictions-and-accuracy-weight-trait-breedings-program-colombian-zebu-brahman.pdf>
22. Ali AK, Shook GE. An Optimun transformation for somatic cell concentration in milk. *J Dairy Sci.* 1980; 63(3):487-490. [https://doi.org/10.3168/jds.s0022-0302\(80\)82959-6](https://doi.org/10.3168/jds.s0022-0302(80)82959-6)
23. 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. <https://doi.org/10.2527/jas.2009-2064>
24. 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. <https://doi.org/10.1186/1753-6561-4-s1-s5>
25. Meuwissen T, Hayes B, Goddard M. Accelerating Improvement of livestock with Genomic Selection. *Annu Rev Anim Biosci.* 2013; 1(1):221-237. <https://doi.org/10.1146/annurev-animal-031412-103705>
26. Chen L, Li C, Zargolzaei M, Schenkel F. Impact of genotypes imputation on the performance of GBLUP and bayesian methods for genomic prediction. *PLoS ONE.* 2014; 9(7):e101544. <https://doi.org/10.1371/journal.pone.0101544>
27. Vázquez AI, Rosa GJ, Weigel KA, de los Campos G, Gianola D, Allison DB. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. *J Dairy Sci.* 2010; 93(12):5942–5949. <https://doi.org/10.3168/jds.2010-3335>
28. Habier D, Rohan LF, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic Selection. *BMC Bioinformatics.* 2011; 12(1):186. <https://doi.org/10.1186/1471-2105-12-186>
29. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci* 2009; 92(2):433–443. <https://doi.org/10.3168/jds.2008-1646> <https://doi.org/pdf/10.4081/ijas.2013.e91>
30. Nicolazzi EL, Negrini R, Chamberlain AJ, Goddard ME, Marsan PA, Hayes BJ. Effect of Prior Distributions on Accuracy of Genomic Breeding Values for Two Dairy Traits. *Ital J Anim Sci* 2013; 12(e91):555-561. <https://www.tandfonline.com/doi/pdf/10.4081/ijas.2013.e91>
31. Colombani C, Legarra A, Fritz S, Guillaume F, Croiseau P, Ducrocq V, et al. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. *J Dairy Sci.* 2013; 96(1):575–591. <https://doi.org/10.3168/jds.2011-5225>

INFORMACIÓN ADICIONAL

;Cómo citar?: Zambrano, J., Echeverri Zuluaga, J., & López Herrera, A. (2019). Evaluación genómica en ganado Holstein Colombiano, usando genotipos imputados a densidad media. Revista MVZ Córdoba, 24(2), 7248-7255. <https://doi.org/10.21897/rmvz.1704>