

Predicción genómica de peso vivo con dos métodos de validación cruzada en ganado bovino

Assessment of genomic prediction of live weight traits using two validation methods in cattle

Francisco Joel Jahuey-Martínez¹, Juan Gabriel Magaña-Monforte², José Candelario Segura-Correa², Juan Carlos Martínez-González³, Raciél J. Estrada-León⁴, Gaspar Manuel Parra-Bracamonte^{5*}

¹Universidad Autónoma de Chihuahua, Facultad de Zootecnia y Ecología. Perif. Francisco. R. Almada Km 1, Pavis Borunda, CP. 31453. Chihuahua, Chihuahua, México

²Universidad Autónoma de Yucatán, Facultad de Medicina Veterinaria y Zootecnia. Carretera Mérida-Xmatkuil Km. 15.5. CP. 97100. Mérida, Yucatán, México.

³Universidad Autónoma de Tamaulipas, Facultad de Ingeniería y Ciencias. Centro Universitario, Campus, CP. 87149. Victoria, Tamaulipas, México.

⁴Tecnológico Nacional de México, Instituto Tecnológico Superior de Calkiní, Avenida AH-Canun, Sin Número, San Felipe, CP. 24900. Calkiní, Campeche, México.

⁵Instituto Politécnico Nacional, Centro de Biotecnología Genómica. Blvd. Del Maestro SN, Esq. Elías Piña, Col. Narciso Mendoza. CP. 88710. Reynosa, Tamaulipas, México.

*Autor de correspondencia: gparra@ipn.mx

Nota científica

Recibida: 08 de junio 2023

Aceptada: 16 de enero 2024

RESUMEN. Se estimó la exactitud de la predicción (EP) de valores genómicos estimados (GEBV) para variables de peso vivo de ganado Charolais utilizando dos métodos de validación cruzada. Se ajustó un modelo BLUP y diferentes métodos de predicción genómica (PG) Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR). La EP fue comparada mediante grupos de validación (GV) formados aleatoriamente y mediante GC. Los resultados mostraron que los tres métodos de PG proporcionaron exactitudes de predicción similares entre los GV pero no exactitudes de predicción superiores a BLUP. La exactitud de predicción de GBLUP y BLUP fue 0.35 y 0.37 para PN, y de 0.30 y 0.41 para PD, respectivamente. Los resultados muestran bajas exactitudes de PG bajo los escenarios evaluados; por lo que para su correcta implementación es necesario incrementar el número de animales y usar valores genéticos desregresados como variables de respuesta.

Palabras clave: BLUP, Charolais, crecimiento, evaluación genómica de un solo paso, SNP, selección genómica.

ABSTRACT. The accuracy of the estimation (AE) of the genomic breeding values (GEBV) of live weight traits of registered Charolais cattle using two cross validation methods was assessed. A BLUP model and different genomic prediction (PG) methods were fitted: Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) and Single-step Bayesian regression (SSBR). AE was compared using randomly formed validation groups (VGs) and CGs. The results showed that all three PG methods provided similar prediction accuracies among GVs but not higher prediction accuracies than BLUP. The prediction accuracy of GBLUP and BLUP was 0.35 and 0.37 for PN, and 0.30 and 0.41 for PD, respectively. The results show that the PG accuracies under the evaluated scenarios are low; Therefore, for its correct implementation, it is necessary to increase the number of animals and use de-returned genetic values as response variables.

Keywords: BLUP, Charolais, genomic selection, single-step genomic evaluation, SNP.

Como citar: Jahuey-Martínez FJ, Magaña-Monforte JG, Segura-Correa JC, Martínez-González JC, Estrada-León RJ, Parra-Bracamonte GM (2024) Predicción genómica de peso vivo con dos métodos de validación cruzada en ganado bovino. Ecosistemas y Recursos Agropecuarios 11(1): e3817. DOI: 10.19136/era.a11n1.3817.

INTRODUCCIÓN

Mediante la predicción genómica, la estimación de valores genómicos (GEBV) en muchas especies domésticas ha sido posible. A diferencia del método de evaluación genética tradicional denominado pedigrí BLUP (PBLUP), el método de predicción genómica, también llamada selección genómica, utiliza la información de marcadores genéticos en lugar de las relaciones del pedigrí para estimar GEBV (Goddard y Hayes 2007, Dekkers *et al.* 2021). Para la implementación ideal de predicción genómica se requiere de una población de referencia conformada por sementales ampliamente probados, donde el efecto de todos los genes sea estimado para una característica de interés, mediante un gran número de polimorfismos de un solo nucleótido (SNP) (VanRaden 2020). La población de predicción necesita estar genéticamente relacionada a la población de referencia. Asumiendo que los marcadores genéticos están en desequilibrio de ligamiento (DL) con los genes causales, los GEBV de los animales candidatos de selección, éstos pueden ser estimados como una función del genotipo individual multiplicado por los efectos previamente estimados de los SNP (VanRaden 2020). Mediante predicción genómica, la selección genética puede realizarse de manera más exacta y a edades más tempranas, provocando una reducción en el intervalo generacional y un incremento en las ganancias genéticas por año, de manera que también se controle la consanguinidad (Meuwissen *et al.* 2001).

Los beneficios de la predicción genómica han sido notificados principalmente en razas de ganado lechero. Por ejemplo, en ganado Holstein de Estados Unidos se ha logrado reducir a menos de la mitad el intervalo generacional y cuadruplicar la ganancia genética para algunas características, esto como consecuencia del uso de los valores de predicción genómica (García-Ruíz *et al.* 2016). La exitosa implementación de la predicción genómica en razas lecheras ha sido posible gracias a las poblaciones de referencia idealmente estructuradas caracterizadas por sementales con un gran número de crías y también debido al buen registro fenotípico y genealógico implementado por los productores (Guinan *et al.* 2023).

Los estudios en ganado de carne también han demostrado que la predicción genómica proporciona mayor exactitud de predicción al estimar GEBV en comparación con el método tradicional PBLUP, incluso cuando las poblaciones de referencia son limitadas en cuanto al número de animales y no están igualmente estructuradas, como en ganado lechero (Misztal *et al.* 2020). Sin embargo, las exactitudes de predicción genómica en ganado de carne no son tan altas como las obtenidas en ganado lechero. Los estudios de predicción genómica en poblaciones de ganado de carne se han enfocado en la exploración para encontrar la mejor estrategia para implementar la predicción genómica investigando, por ejemplo, diferentes modelos estadísticos de predicción (Lee *et al.* 2017), densidades de marcadores (Gunia *et al.* 2014), relaciones entre la población de referencia y la población de predicción (Saatchi *et al.* 2012), predicciones uni o multirraciales (Neves *et al.* 2014), entre otros aspectos que impactan la exactitud de predicción.

Como uno de los mayores productores de ganado de carne a nivel mundial (USDA 2023), México podría incrementar su producción y la rentabilidad en los sistemas de producción de carne, a través de la selección genética para mejorar las características productivas del ganado. Sin embargo, el mejoramiento genético de características productivas en algunas razas bovinas para carne ha sido limitado, por ejemplo, en ganado Charolais (Parra-Bracamonte *et al.* 2016). Por lo tanto, el objetivo de este estudio fue investigar el impacto de la información genómica en la

exactitud de predicción de GEBV de dos características de peso vivo de ganado Charolais de registro utilizando dos métodos de validación cruzada.

MATERIALES Y MÉTODOS

Datos de estudio

La población de estudio consistió de 823 bovinos Charolais (683 hembras y 140 machos) previamente genotipificados con el panel GeneSeek Genomic Profiler Bovine HD de 77 mil SNP (GGPHD, Neogen Corp., Lincoln, NE) (Jahuey-Martínez *et al.* 2016). Esta población pertenecía a tres hatos del norte de México y representó la progenie de 138 sementales y 577 madres. Los registros fenotípicos de peso al nacimiento (PN) y peso al destete ajustado a los 205 días (PD) de la población genotipificada, así como, los registros fenotípicos de 5 575 animales no genotipificados fueron proporcionados por la Asociación Charolais Charbray Herd Book de México. Estos animales y sus ancestros conformaron un pedigrí de 6 generaciones que incluyó 7 256 animales.

Control de calidad de los datos genotípicos

Los datos genotípicos fueron previamente evaluados para descartar SNP y muestras cuya calidad fuera pobre (Jahuey-Martínez *et al.* 2016). Los parámetros de calidad para los SNP fueron: que los SNPs fueran polimórficos, una tasa de llamado > 90%, frecuencia del alelo menor > 0.01, la no desviación al equilibrio de Hardy-Weinberg (valor de $P > 1 \times 10^{-15}$), y la localización autosómica y con posición definida en el ensamblaje del genoma bovino UMD versión 3.1. Los parámetros de calidad para las muestras fueron: tasa de llamado > 80%, heterocigocidad < a 3 DS y correlación genotípica < 0.98. Un total de 68,337 SNP cumplieron con los parámetros de calidad y fueron utilizados en los análisis de predicción genómica. El control de calidad se realizó con el software SNPQC (Rutkoski *et al.* 2013) de R (R Core Team 2017).

Los genotipos faltantes fueron reemplazados por el valor promedio observado para cada SNP. El grado de DL fue investigado calculando el estadístico r^2 entre cada par de SNP separados por diferentes distancias genómicas.

Variables de respuesta

Se calcularon los fenotipos ajustados (\hat{y}) por efectos ambientales para ser utilizados como variables de respuesta en los análisis de predicción genómica. Los \hat{y} fueron obtenidos como los residuales de un modelo lineal que, que incluyeron los efectos fijos de grupo contemporáneo (GC) y edad de la madre. Los GC fueron formados como la combinación de las categorías del sexo, hato, año y época de nacimiento del animal. En la Tabla 1 se muestran los estadísticos descriptivos de las características fenotípicas y las variables de respuesta utilizadas en este estudio. Los fenotipos ajustados fueron utilizados como variables de respuesta en lugar de EBV o EBV desregresados debido a que la mayoría de los animales en este estudio fueron hembras y en consecuencia tuvieron exactitudes bajas de EBV.

Tabla 1. Descripción estadística de los fenotipos y fenotipos ajustados de peso vivo de ganado Charolais (kg).

| Variable ¹ | N | Media | DE | Mediana | Min ² | Max ³ | EE |
|-----------------------|-------|--------|-------|---------|------------------|------------------|------|
| PN | 6,195 | 40.51 | 7.39 | 40 | 15 | 66 | 0.09 |
| PD | 5,727 | 228.64 | 41.83 | 226.9 | 101.47 | 396.59 | 0.55 |
| PN* | 6,195 | 0 | 5.87 | 0 | -28.9 | 21.74 | 0.07 |
| PD* | 5,727 | 0 | 30.64 | -0.81 | -141.82 | 159.64 | 0.4 |

¹PN, PD, and * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente. DE = desviación estándar, EE = error estándar; ²Min = valor mínimo; ³Max = valor máximo.

Predicción de EBV basada en el pedigrí (BLUP)

El análisis de BLUP fue realizado utilizando los datos fenotípicos de animales tanto genotipificados como no genotipificados. El modelo general utilizado en PBLUP fue el siguiente:

$$y = 1\mu + Zu + e,$$

Donde: y es el vector de fenotipos corregidos (\hat{y}), 1 es un vector de 1s, μ es el intercepto, Z es la matriz de diseño que asigna los registros a los efectos genético-aditivos, u es el vector de efectos genético aditivos y e es el vector aleatorio de residuales. En este modelo se asume que $u \sim N(0, A\sigma_g^2)$, donde A es la matriz de relaciones del numerador derivada del pedigrí y σ_g^2 es la varianza genética aditiva. Se asumió que los residuales se distribuían idéntica e independientemente con media cero y varianza σ_e^2 . Los componentes de la varianza y las estimaciones de heredabilidad (h^2) se calcularon a partir de los datos fenotípicos crudos y utilizando el modelo anterior más los efectos fijos antes mencionados. Estos análisis fueron realizados utilizando REML en el paquete breedR del programa de software R (R Core Team 2021).

Métodos de predicción genómica

Se evaluaron los métodos Genomic-based best linear unbiased prediction (GBLUP) y Bayes C (BC) utilizando únicamente los datos de la población genotipificada. El modelo de GBLUP fue similar al utilizado en BLUP, pero una matriz de relaciones genómicas (G) estimadas de los marcadores genéticos fue utilizada en lugar de la matriz A . La matriz G fue estandarizada para tener un valor diagonal de aproximadamente uno seguido de una descomposición de autovectores. En el modelo bayesiano, Zu fue remplazado por Wg donde W es la matriz de genotipos codificada como 0, 1 y 2 representando el número de alelos B para cada marcador y para cada animal y g es el vector efectos de los marcadores. Los métodos GBLUP y BC fueron evaluados en el paquete BGLR (Pérez y de los Campos 2014) de R (R Core Team 2021) utilizando los parámetros por defecto para cada modelo corriendo 50 000 iteraciones e incluyendo un periodo de descarte (burnin) de 5 000 muestras. En BGLR, el supuesto a priori para los efectos de los SNP en el modelo BC fue una mezcla de una distribución con mayor masa de probabilidad en el entorno del cero y una distribución normal (Habier *et al.* 2011). Los resultados de las cadenas de iteraciones se guardaron y exportaron en archivos, por lo que las medias posteriores se calcularon con 45 000 iteraciones (50 000-5 000).

Adicionalmente, con el objetivo de combinar la información de animales genotipificados y no genotipificados en un modelo de predicción genómica, se evaluó el método Single-step

Bayesian Regression (SSBR; Fernando *et al.* 2014). Similar al método Single step GBLUP (Legarra *et al.* 2009), SSBR permite incorporar la matriz de relaciones genómicas derivada de los SNP en la matriz de relaciones genéticas derivada del pedigrí y de esta manera, la matriz de relaciones combinada es utilizada para calcular GEBVs para todos los animales en el pedigrí (tanto genotificados como no genotificados). En el modelo SSBR, se modelaron los fenotipos ajustados de acuerdo con Lee *et al.* (2017):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{ZM}\boldsymbol{\alpha} + \mathbf{Z}_n\boldsymbol{\epsilon} + \mathbf{e}.$$

Donde: \mathbf{y} es el vector de fenotipos ajustados para los animales genotificados y no genotificados, $\mathbf{X} = \begin{bmatrix} \mathbf{1} & -\mathbf{Z}_n\mathbf{A}_{ng}\mathbf{A}_{gg}^{-1}\mathbf{1} \\ \mathbf{1} & -\mathbf{Z}_g\mathbf{1} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \mu_g \end{bmatrix}$, μ es la media general, μ_g representa la diferencia en los efectos genético aditivos entre los animales genotificados y no genotificados, \mathbf{Z} es la matriz de diseño, $\mathbf{M} = \begin{bmatrix} \widehat{\mathbf{M}}_n \\ \mathbf{M}_g \end{bmatrix}$, donde \mathbf{M}_g es la matriz de covariables SNP para los animales genotificados y $\widehat{\mathbf{M}}_n = \mathbf{A}_{ng}\mathbf{A}_{gg}^{-1}\mathbf{M}_g$, representa las covariables SNP imputadas para los animales no genotificados que son derivados de sus parientes genotificados, $\boldsymbol{\epsilon}$ es el residual de imputación, \mathbf{Z}_n y \mathbf{Z}_g son las matrices de diseño asignando registros a los efectos genético aditivos de los animales no genotificados y genotificados, \mathbf{A}_{ug} es la partición de \mathbf{A} que corresponde a las relaciones entre los animales genotificados y no genotificados y \mathbf{A}_{gg} es la partición de \mathbf{A} que corresponde a los animales genotificados. Los supuestos para μ , μ_g , \mathbf{e} y $\boldsymbol{\epsilon}$ fueron los mismos que los descritos por Lee *et al.* (2017). Para este método, se utilizó el modelo BC asumiendo que una proporción de 0.995 de los SNP no tenían efecto sobre la característica bajo estudio. El método SSBR se evaluó utilizando 50,000 iteraciones en el paquete JWAS (2017) en lenguaje de programación Julia.

Validación cruzada

Se evaluaron dos escenarios de poblaciones de validación (PV) basados en la elección de animales aleatoriamente (PV_AL) y de acuerdo con su GC (PV_GC). Las PV_AL fueron creadas seleccionando 82-83 animales de la población genotificada hasta que se obtuvieran 10 PV_AL. Las PV_GC fueron creadas seleccionando todos los animales pertenecientes a GC que previamente habían sido clasificados por su tamaño (número de animales). Para clasificar los GC, se definieron tres tamaños: GC pequeños, medianos y grandes los cuales incluían 1-6, 7-14 y 15-49 animales, respectivamente. Como una breve descripción, la clase más abundante en la presente base de datos fue la clase pequeña con 105 GC y 270 animales, seguida por la clase mediana con 26 GC y 259 animales y finalmente la clase grande con 13 GC y 294 animales. Al final, se seleccionaron 10, 2 y 1 GC de las clases pequeña, mediana y grande, respectivamente y crearon 12 PV_GC las cuales contenían en promedio 69 animales.

La exactitud de predicción de cada método fue calculada como la correlación entre los GEBV y los fenotipos ajustados de los animales en el conjunto de validación dividido por la raíz cuadrada de la h^2 de la característica, según Meuwissen *et al.* (2013). También, se calcularon el coeficiente de regresión (CR) del fenotipo ajustado sobre el GEBV y el error cuadrático medio (ECM).

RESULTADOS Y DISCUSIÓN

En la Tabla 2 describe las estimaciones de los componentes de la varianza y heredabilidad de pesos vivos de ganado Charolais utilizados. La heredabilidad para el PN fue 0.34 y para PD 0.22.

Tabla 2. Estimaciones de los componentes de la varianza¹ y heredabilidad (h^2) de peso vivo (kg).

| Variable ² | σ^2_g | σ^2_e | h^2 |
|-----------------------|--------------|--------------|-------|
| PN | 13.213 | 25.254 | 0.343 |
| PD | 232.78 | 620.92 | 0.229 |

¹Los componentes de varianza son la varianza genética aditiva (σ^2_g) y la varianza ambiental (σ^2_e). ² PN y PD son el peso al nacimiento y al destete.

En tanto que la Figura 1, muestra el decaimiento de DL sobre las distancias genómicas de los marcadores utilizados en la estimación de los GEBV.

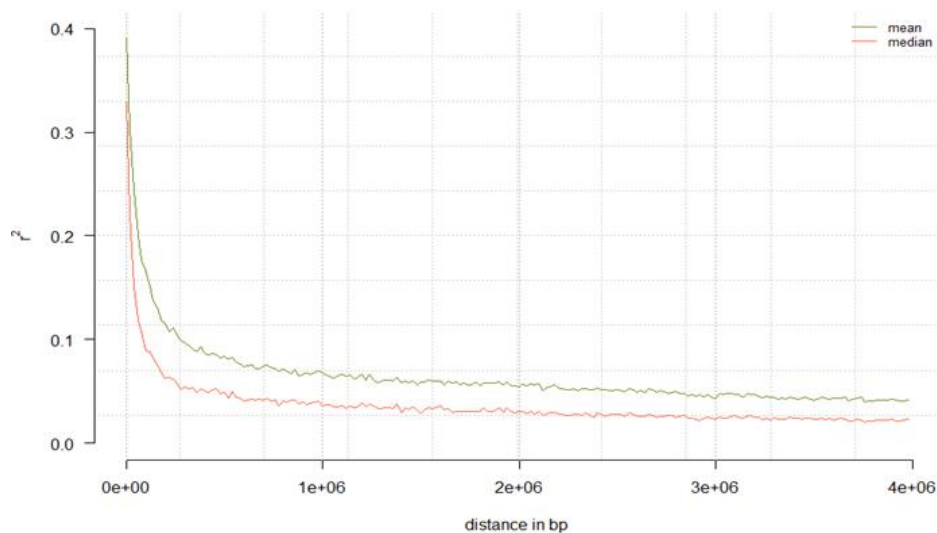


Figura 1. Descenso del desequilibrio de ligamiento (r^2) sobre varias distancias genómicas del arreglo utilizado para las predicciones.

El fundamento de la predicción genómica se basa en que el efecto de todos los genes sobre una característica se puede estimar de manera aproximada en una población de referencia mediante un gran número de marcadores SNP en DL con los genes causales. Como primer paso, en este estudio se caracterizó el DL a partir de los marcadores SNP. Los valores de r^2 promedio entre pares de SNP en distancias cromosómicas de 0-25 kilo pb, 25-50 kilo pb, 50-100 kilo pb, 100-500 kilo pb, 0.5-1 mega pb, 1-5 mega pb y 5-10 mega pb fueron 0.36, 0.27, 0.19, 0.10, 0.07, 0.04 y 0.02, respectivamente. El DL observado en este estudio fue similar a lo reportado en una población de ganado Charolais de Canadá (Lu *et al.* 2012), así como en otras razas bovinas (Zhu *et al.* 2013).

La predicción fue moderada para todos los métodos de predicción genómica y similares entre SSBP y BLUP, para los dos métodos de validación cruzada (Tabla 3). Las exactitudes de predicción obtenidas fueron similares entre los modelos de predicción genómica y las clases de PV

(Tabla 3). Contrario a otros estudios, las exactitudes de predicción obtenida mediante los modelos de predicción genómica no fueron mayores a las exactitudes de predicción obtenidas mediante BLUP. Por ejemplo, la exactitud de predicción obtenida mediante GBLUP fue 0.35 y 0.30 mientras que la exactitud de predicción mediante BLUP fue 0.37 y 0.41 para PN y PD, respectivamente.

Tabla 3. Exactitud de predicción de varios métodos de predicción genómica¹ versus BLUP con dos métodos de validación cruzada.

| Variable ² | PV | GBLUP | BC | SSBR | BLUP |
|-----------------------|----|-------|-------|-------|-------|
| PN* | AL | 0.312 | 0.311 | 0.379 | 0.381 |
| | GC | 0.356 | 0.353 | 0.371 | 0.372 |
| PD* | AL | 0.290 | 0.295 | 0.345 | 0.393 |
| | GC | 0.307 | 0.310 | 0.370 | 0.414 |

¹Los métodos de predicción genómica son Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR). ²PN, PD y * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente. ³PV = tipo de población de validación: AL: Aleatoria GC: Por grupo contemporáneo.

Se ha demostrado que los métodos de predicción de GEBV basados en el método de single-step, que toman en cuenta la información de animales no genotificados, predicen con mayor exactitud que los modelos bayesianos y GBLUP (Lee *et al.* 2017). Por esta razón, en este estudio también se evaluó el método basado en single-step denominado SSBR esperando obtener mayores exactitudes de predicción. Sin embargo, la inclusión de datos fenotípicos y de pedigrí de animales no genotificados no tuvo efecto sobre la exactitud de predicción. La exactitud de predicción obtenida mediante SSBR fue 0.37 tanto para PN como para PD.

Estudios en otras razas bovinas de carne han reportado mayores exactitudes de predicción y demuestran que mediante predicción genómica se pueden obtener estimaciones más exactas del mérito genético en comparación con BLUP. Las exactitudes promedio de GEBV para PN y PD en las razas Angus, Hereford, Charolais, Limousin y Simmental fueron 0.70 y 0.53 (Saatchi *et al.* 2011), 0.42 y 0.21 (Saatchi *et al.* 2013), 0.42 y 0.45 (Gunia *et al.* 2014), 0.58 y 0.58 (Saatchi *et al.* 2012), 0.65 y 0.52 (Saatchi *et al.* 2012), respectivamente. Recientemente, Peters *et al.* (2023) en una población de novillas de la raza Brangus con diferentes métodos de PG, reportaron exactitudes entre 0.640 y 0.402 con clustering de k-medias, y 0.823 y 0.500 con clustering aleatorio para PN, y valores entre 0.425 y 0.220 con clustering de k-medias, y 0.768 y 0.386 con clustering aleatorio para PD. Es importante destacar que en esos estudios se utilizaron mayor número de animales (principalmente sementales) en los análisis de predicción. Además, los GEBV fueron estimados a partir de EBV desregresados, en lugar de los fenotipos observados. En México, el primer estudio de predicción genómica utilizando una población de ganado Charolais, fue publicado por Jahuey-Martínez *et al.* (2020) quienes encontraron valores similares de baja exactitud con un método de análisis en BLUP y de validación diferente al del presente estudio, encontrando ventajas para el método de predicción SSBR utilizando el método de validación por GC.

Para las estimaciones de los coeficientes de regresión se observaron mayores valores para BC y GBLUP comparados con SSBR y BLUP, que fueron similares. Por su parte, los valores estimados de error cuadrático fueron muy similares entre modelos, independientemente del método de validación cruzada (Tabla 4). El valor de error de SSBR fue evidentemente menor con respecto a los otros métodos estudiados por grupo contemporáneo.

Tabla 4. Estimaciones de los coeficientes de regresión y error cuadrático medio de BLUP y métodos de predicción genómica¹.

| Variable ⁴ | PV ⁵ | CR ² | | | | ECM ³ | | | |
|-----------------------|-----------------|-----------------|------|------|------|------------------|--------|--------|--------|
| | | GBLUP | BC | SSBR | BLUP | GBLUP | BC | SSBR | BLUP |
| PN* | AL | 0.82 | 0.84 | 0.81 | 0.75 | 28.06 | 27.97 | 27.79 | 27.68 |
| PN* | GC | 0.85 | 0.89 | 0.76 | 0.71 | 26.87 | 26.88 | 26.90 | 26.82 |
| PD* | AL | 0.80 | 0.84 | 0.67 | 0.72 | 681.31 | 679.18 | 102.49 | 675.41 |
| PD* | GC | 0.84 | 0.87 | 0.78 | 0.81 | 674.77 | 673.62 | 99.90 | 669.34 |

¹Los métodos de predicción genómica son Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR). ²RC = coeficiente de regresión. ³MSE = error cuadrático medio. ⁴PN, PD y * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente. ⁵VS = tipo de población de validación: AL: Aleatoria GC: Por grupo contemporáneo.

La predicción en el presente estudio pudo haber sido influida por factores como el tipo de variable de respuesta y el tamaño de la población, los cuales son de los principales factores que impactan la predicción genómica. Los fenotipos corregidos como variables de respuesta no fueron la mejor opción para usar en predicción genómica debido a que estos fenotipos aún permanecen influenciados por factores no genéticos. En este estudio, las dos características de peso corporal sólo fueron corregidas por los efectos de GC y edad de la madre. En el modelo de ajuste podrían considerarse otros factores como el efecto permanente de la madre para obtener fenotipos mejor corregidos. Sin embargo, los datos utilizados no contenían suficiente información para considerar ese efecto. Además, en el modelo de ajuste el número de observaciones por nivel en cada factor ambiental fue bajo como el caso del efecto de GC. Considerando únicamente los datos de la población genotificada, se observaron más de 100 niveles de GC con menos de seis observaciones. A pesar de esto, la validación mediante GC, mostró menor error cuadrático medio con el método de VC, cuando el método de predicción fue SSBR, lo que apoya las ventajas del método. Otros estudios han reportado ventajas de métodos similares de predicción genómica de un solo paso, entre ellas la disminución del error cuadrático medio (Naserkheil *et al.*, 2020).

Otro factor limitante en este estudio fue que los animales genotificados eran en su mayoría hembras, las cuales fueron seleccionadas con base en la participación de los productores y a la disponibilidad de datos fenotípicos, en lugar de realizar una selección de sementales con alta exactitud de EBV. El uso de poblaciones de referencia representadas por sementales altamente probados sería una mejor opción para investigar el impacto de la información genómica en los modelos de predicción genómica. En México, el número total de sementales de la raza Charolais con 10 o más crías ha ascendido a 4 407, de acuerdo con la última evaluación genética de esta raza (Asociación Charolais-Charbray Herdbook de México, 2020). En el futuro, la construcción de una adecuada población de referencia debería necesariamente considerar la identificación de

sementales altamente probados y la inclusión de un mayor número de productores en un programa de selección genómica de mayor magnitud.

A pesar de que se han realizado algunos estudios a nivel local que demuestran la importancia del registro de pedigrís, el impacto de usar EBV como herramienta de selección (Parra-Bracamonte *et al.* 2016) y el potencial del uso de marcadores genéticos para mejorar características productivas, los resultados de estas investigaciones aún no han sido puestos en práctica. Desafortunadamente, la cultura de cría de ganado en México todavía sigue siendo dominada por la selección de animales con base en aspectos fenotípicos observables y esquemas de reproducción que no toman en cuenta la importancia genética de los animales. Por lo tanto, se requiere de la implementación de esquemas de reproducción y técnicas de selección efectivas que utilicen o no información genómica. Para esto, una tarea de los investigadores sería implementar cursos de capacitación que familiaricen a los productores en los conceptos generales de la selección animal y proporcionar programas estadísticos útiles para el registro, manejo y simulación de procedimientos donde los productores puedan observar fácilmente estadísticas de sus hatos e indicadores acerca de posibles cambios genéticos, de acuerdo con ciertas condiciones de selección y manejo. De esta manera, los productores podrían reconocer la importancia del correcto registro de pedigrís e información fenotípica e incluso la importancia de la información genómica.

El registro de características productivas en ganado de carne es importante en términos de realizar predicción genómica. La predicción genómica no puede ser utilizada si no existen los datos fenotípicos en la población de referencia y no habría ninguna ventaja de genotipificar animales. Tampoco existe un objetivo de cría, así como los correspondientes registros fenotípicos. Para la raza Charolais, solo algunas características productivas (peso vivo, talla, circunferencia escrotal, permanencia productiva, caracteres de canal por ultrasonido) son consideradas para evaluación genética (Asociación Charolais-Charbray Herdbook de México, 2020). Por lo tanto, sólo para estas características sería posible estimar GEBV si se implementara la predicción genómica en esta raza. Desafortunadamente, muchas características importantes que impactan en la producción; por ejemplo, eficiencia alimenticia, temperamento, resistencia a parásitos o a diversos climas, no son registradas en las razas de ganado bovino para carne en México y es precisamente en este tipo de características donde la predicción genómica puede ser de mayor beneficio, ya que, son características de baja heredabilidad y difíciles o costosas de medir. Otras características fenotípicas podrían ser consideradas en las razas bovinas para carne, pero estas necesitarían estar relacionadas con los objetivos de la mejora genética.

El costo de genotipificación es un factor determinante en la implementación de la predicción genómica debido a la alta inversión económica que representa para los productores. En México, los servicios de genotipificación de especies de importancia doméstica son provistos por compañías del extranjero, y aunque los costos del servicio cada vez son menores, el tipo de cambio monetario sigue siendo una limitante importante. Experiencias y estrategias de otros proyectos genómicos podrían ser consideradas en futuras aplicaciones de predicción genómica para ayudar a minimizar el costo de genotipificación y maximizar la exactitud de predicción. Por ejemplo, la imputación de genotipos de animales no genotipificados podría realizarse para incrementar el tamaño de la población de referencia (Pimmentel *et al.* 2013, Da-Silva *et al.* 2023). Además, el genotipo de sementales probados por progenie puede ser predicho de manera exacta mediante la estructuración de una base de datos que incluya el genotipo de familias de medios hermanos y

mediante el uso de programas de imputación como hspbase (Ferdosi *et al.* 2014). Siguiendo esta estrategia, fue posible predecir el genotipo de 66 sementales no genotificados a partir de la información genotípica de 636 animales genotificados. En un análisis complementario previo a este estudio, se observó que el genotipo del semental puede ser imputado de manera exacta hasta con 66K SNP teniendo al menos 15 crías genotificadas por semental. Otras estrategias como la imputación de baja a alta densidad de SNP también podrían implementarse (Huang *et al.* 2012).

Se demostró que es posible la implementación de métodos de predicción genómica para características de peso vivo en ganado Charolais, pero sin ningún incremento en la exactitud de predicción del mérito genético para estas características. La principal razón de esta falta de mejora fue debido al tamaño reducido de la población de referencia y al uso de fenotipos ajustados en vez de valores genéticos desregresados de sementales con exactitudes altas o gran número de proge. Se esperaría que en un futuro se pueda tener acceso a la información de EBV y exactitudes recientes para evaluar EBV desregresados como variables de respuesta y considerar los genotipos imputados de sementales en los modelos de predicción genómica.

AGRADECIMIENTOS

Los autores desean expresar su agradecimiento a la Asociación Charolais Charbray Herd Book de México, por las facilidades prestadas para la realización del presente estudio.

CONFLICTO DE INTERÉS

Los autores declaran que no tienen intereses en competencia.

LITERATURA CITADA

- Asociación Charolais-Charbray Herdbook de México (2020) Sumario de sementales Charolais y Charbray. <https://charolais.org.mx/wp-content/uploads/2020/08/Sumario-CHAROLAIS-2020-digital-B.pdf>. Fecha de consulta: 15 de marzo de 2023.
- Da-Silva NJB, Peripoli E, Pereira AS, Stafuzza NB, Lôbo RB, Fukumasu H, Ferraz JBS (2023) Weighted genomic prediction for growth and carcass-related traits in Nelore cattle. *Animal Genetics* 54(3): 271-283. <https://doi.org/10.1111/age.13310>
- Dekkers JC, Su H, Cheng J (2021) Predicting the accuracy of genomic predictions. *Genetics Selection Evolution* 53: 1-23. <https://doi.org/10.1186/s12711-021-00647-w>
- Ferdosi MH, Kinghorn BP, van der Werf JH, Lee SH, Gondro C (2014) Hspbase: An R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. *BMC Bioinformatics* 15: 172. <https://doi.org/10.1186/1471-2105-15-172>
- Fernando RL, Dekkers JCM, Garrick DJ (2014) A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetic Selection Evolution* 46: 50. <https://doi.org/10.1186/1297-9686-46-50>.
- García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences U S A.* 113(28): E3995-4004. <https://doi.org/10.1073/pnas.1519061113>

- Goddard ME, Hayes BJ (2007) Genomic selection. *Journal of Animal breeding and Genetics* 124(6): 323-330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Guinan FL, Wiggans GR, Norman HD, Dürr JW, Cole JB, Van Tassell CP, Misztal I, Lourenco D. (2023) Changes in genetic trends in US dairy cattle since the implementation of genomic selection. *Journal of Dairy Science* 106(2): 1110-1129. <https://doi.org/10.3168/jds.2022-22205>
- Gunia M, Saintilan R, Venot E, Hozé C, Fouilloux MN, Phocas F (2014) Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science* 92(8): 3258-3269. <https://doi.org/10.2527/jas.2013-7478>
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12: 186. <https://doi.org/10.1186/1471-2105-12-186>
- Huang Y, Hickey JM, Cleveland MA, Maltecca C (2012) Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genetic Selection Evolution* 44: 25. <https://doi.org/10.1186/1297-9686-44-25>
- Jahuey-Martínez FJ, Parra-Bracamonte GM, Sifuentes-Rincón AM, Martínez-González JC, Gondro C, García-Pérez CA, López-Bustamante LA (2016) Genomewide association analysis of growth traits in Charolais beef cattle. *Journal of Animal Science* 94(11): 4570-4582. <https://doi.org/10.2527/jas.2016-0359>
- Jahuey-Martínez FJ, Parra-Bracamonte GM, Garrick DJ, López-Villalobos N, Martínez-González JC, Sifuentes-Rincón AM, López-Bustamante LA (2020) Accuracies of direct genomic breeding values for birth and weaning weights of registered Charolais cattle in Mexico. *Animal Production Science* 60(6): 772-779. <https://doi.org/10.1071/AN18363>
- JWAS (2017) Julia implementation of whole-genome analyses software using univariate and multivariate Bayesian mixed effects model. <http://QTL.rocks>. Fecha de consulta: 20 de febrero de 2023.
- Lee J, Cheng H, Garrick D, Golden B, Dekkers J, Park K, Lee D, Fernando R (2017) Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. *Genetics Selection Evolution* 49(1): 2. <https://doi.org/10.1186/s12711-016-0279-9>
- Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92: 4656-4663. <https://doi.org/10.3168/jds.2009-2061>
- Meuwissen T, Hayes B, Goddard M (2013) Accelerating improvement of livestock with genomic selection. *Annual Reviews of Animal Biosciences* 1: 221-37. <https://doi.org/10.1146/annurev-animal-031412-103705>
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4): 1819-1829.
- Misztal I, Lourenco D, Legarra A (2020) Current status of genomic evaluation. *Journal of Animal Science* 98(4): 1-14. <https://doi.org/10.1093/jas/skaa101>
- Naserkheil M, Lee DH, Mehrban H (2020) Improving the accuracy of genomic evaluation for linear body measurement traits using single-step genomic best linear unbiased prediction in Hanwoo beef cattle. *BMC genetics* 21(1): 1-9. <https://doi.org/10.1186/s12863-020-00928-1>
- Parra-Bracamonte GM, Lopez-Villalobos N, Morris ST, Sifuentes-Rincón AM, Lopez-Bustamante LA (2016) Genetic trends for live weight traits reflect breeding strategies in registered Charolais Farms in Mexico. *Tropical Animal Health and Production* 48(8): 1729-1738. <https://doi.org/10.1007/s11250-016-1150-2>
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198(2): 483-495. <https://doi.org/10.1534/genetics.114.164442>
- Peters SO, Kizilkaya K, Sinecen M, Mestav B, Thiruvenkadan AK, Thomas MG (2023) Genomic Prediction Accuracies for Growth and Carcass Traits in a Brangus Heifer Population. *Animals* 13(7): 1272. <https://doi.org/10.3390/ani13071272>
- Pimentel EC, Wensch-Dorendorf M, König S, Swalve HH (2013) Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genetics Selection Evolution* 45: 12. <https://doi.org/10.1186/1297-9686-45-12>
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Fecha de consulta: 17 de enero de 2020.
- Rutkoski JE, Poland J, Jannink JL, Sorrells ME (2013) Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes | Genomes | Genetics* 3(3): 427-439. <https://doi.org/10.1534/g3.112.005363>

- Saatchi M, Ward J, Garrick DJ (2013) Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *Journal of Animal Science* 91(4): 1538-1551. <https://doi.org/10.2527/jas.2012-5593>
- Saatchi M, McClure MC, McKay SD, Rolf MD, Kim J, Decker JE, Tasia M, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RD, Garrick DJ, Taylor JF (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution* 43(1): 40. <https://doi.org/10.1186/1297-9686-43-40>
- Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ (2012) Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genetics Selection Evolution* 44(1): 38. <https://doi.org/10.1186/1297-9686-44-38>
- USDA (2023) Livestock and Poultry: World Markets and Trade. https://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf. Fecha de consulta: 11 de abril de 2023.
- VanRaden PM (2020) Symposium review: How to implement genomic selection. *Journal of Dairy Science* 103(6): 5291-5301. <https://doi.org/10.3168/jds.2019-17684>