



Minor allele frequency in genomic prediction for growth traits in Braunvieh cattle

Frecuencia del alelo menor en predicción genómica para características de crecimiento en bovinos Suizo Europeo

Trujano-Chavez, M. Z., Valerio-Hernández, J. E., López-Ordaz, R., Ruíz-Flores, A.

Universidad Autónoma Chapingo, Posgrado en Producción Animal,
Departamento de Zootecnia, Estado de México, México.

Cite this paper/Como citar este artículo: Trujano-Chavez, M. Z., Valerio-Hernández, J. E., López-Ordaz, R., Ruíz-Flores, A. (2021). Minor allele frequency in genomic prediction for growth traits in Braunvieh cattle. *Revista Bio Ciencias* 8, e1052. doi: <https://doi.org/10.15741/revbio.08.e1052>



ABSTRACT

The present research aimed to compare the effect of minor allele frequency (MAF) threshold during the quality control process of genotypes, using single nucleotide polymorphisms (SNPs) to predict genomic values of growth traits in Braunvieh cattle. The analysis was performed including 28,973 and 18,994 phenotypic records of birth and weaning weights, 12,835 single nucleotide polymorphisms and 300 animals. The traits were analyzed with the single-step genomic evaluation approach. The criteria for comparison of the minor allele frequency thresholds compared were: 1) ranking of genomic values: Spearman and Pearson correlation coefficient estimates were obtained for genomic values predicted with standard (0.05) and lower levels of minor allele frequency; 2) prediction ability of models: cross-

Article Info/Información del artículo

Received/Recibido: August 29th 2020.

Accepted/Aceptado: March 22th 2021.

Available on line/Publicado: April 17th 2021.

RESUMEN

En el estudio se compararon umbrales de frecuencia del alelo menor (MAF) durante el control de calidad de los genotipos, utilizando polimorfismos de nucleótido simple (SNPs) para predecir valores genómicos de características de crecimiento en bovinos Suizo Europeo. El análisis se realizó con 28,973 y 18,994 registros de pesos al nacimiento y destete, 12,835 polimorfismos de nucleótido simple y 300 animales. Las características se analizaron bajo el enfoque de evaluación genómica en un solo paso. Los criterios de comparación fueron: 1) jerarquización de valores genómicos: se estimaron coeficientes de correlación Spearman y Pearson entre valores genómicos predichos con frecuencia estándar del alelo menor (0.05) y niveles menores; 2) habilidad de predicción de los modelos: se realizó validación cruzada con cuatro repeticiones y se obtuvo el promedio de los coeficientes de correlación de los fenotipos reales contra predichos; y 3) coeficiente de regresión simple: los valores genómicos obtenidos con frecuencia del alelo menor estándar e inferiores, fueron las variables independiente y dependiente, respectivamente. El efecto del

*Corresponding Author:

Agustín Ruíz-Flores. Universidad Autónoma Chapingo, Carretera Federal México-Texcoco Km 38.5, C.P. 56227, Texcoco, Estado de México, México. Phone: (595) 952 1683. E-mail: aruizf@chapingo.mx, <https://produccionanimal.chapingo.mx/>

validation with four replicates was performed, average of correlation between real and predicted phenotypes was obtained; and 3) regression coefficient estimates: the genomic values obtained with standard and lower levels of minor allele frequency were used as independent and dependent variables, respectively. The effect of minor allele frequency threshold on the ranking of genomic values was not important (correlation estimates greater than 0.99, $<2.2 \times 10^{-16}$). The prediction ability model was similar: around 0.7. the regression coefficient estimates were close to one (0.99, $<2 \times 10^{-16}$). The results suggest that a minor allele frequency between 0.0 and 0.05, does not influence the genomic value prediction in small populations. However, these results should be regarded as preliminary and susceptible to changes if the study is repeated with a greater number of markers and animals.

KEY WORDS

Single nucleotide polymorphism, minor allele frequency, genotype quality control, prediction of genetic values, Braunvieh.

Introduction

The discovery of large numbers of single nucleotide polymorphisms (SNPs) markers; the availability of high throughput technology to genotype animals at thousands of SNPs in a cost-effective manner and advances in statistical methodologies for analysis, helped to generalize the use of genomic selection (GS) in the world (Meuwissen *et al.*, 2016). Currently, the additive effect of molecular markers in the prediction of genetic values (GV) of traits of economic interest in animals is being studied (De los Campos *et al.*, 2013).

The success of exploiting SNP markers in GS lies in achieving accurate predictions of GVs for the traits studied. According to Meuwissen *et al.* (2016) individuals with reliable phenotypic information and as many SNPs as possible are required. To improve the accuracy of GV predictions, variants of the innovative methodology of (Meuwissen *et al.*, 2001), such as semi-parametric (Gianola *et al.*, 2006) or non-parametric (Gianola *et al.*, 2009) approaches, have been proposed. In addition, a wide number of studies on methodological modifications and contributions have been proposed, such as the use of K-means clustering for cross-

nivel de frecuencia del alelo menor en la jerarquización de los valores genómicos no fue importante (correlaciones mayores que 0.99, $<2.2 \times 10^{-16}$). La habilidad predictiva del modelo fue similar: 0.7 promedio. Los coeficientes de regresión fueron cercanos a uno (0.99, $<2 \times 10^{-16}$). Los resultados sugieren que una frecuencia del alelo menor entre 0.0 y 0.05, no altera las predicciones de valores genómicos en poblaciones pequeñas. Sin embargo, estos resultados son preliminares y susceptibles a cambios si el estudio se repite con un número de marcadores y animales mayor.

PALABRAS CLAVE

Polimorfismo de nucleótido simple, frecuencia del alelo menor, control de calidad de genotipos, predicción de valores genómicos, Pardo Suizo.

Introducción

El descubrimiento de cantidades grandes de marcadores polimorfismos de un solo nucleótido (SNPs); la tecnología de rendimiento alto para genotipar animales en miles de SNPs de manera rentable y los avances en metodologías estadísticas para análisis, ayudaron a generalizar el uso de selección genómica (SG) en el mundo (Meuwissen *et al.*, 2016). Actualmente se estudia el efecto en conjunto de marcadores moleculares en la predicción de valores genéticos (VG) de características de interés económico en los animales (De los Campos *et al.*, 2013).

El éxito del aprovechamiento de marcadores SNP en SG reside en lograr predicciones precisas de los VG para las características estudiadas. Según Meuwissen *et al.* (2016) se requieren individuos con información fenotípica confiable y con la mayor cantidad posible de SNPs. Para mejorar la precisión de las predicciones de los VG, se han propuesto variantes de la metodología innovadora de (Meuwissen *et al.*, 2001), como los enfoques semiparamétricos (Gianola *et al.*, 2006) o no paramétricos (Gianola *et al.*, 2009). Además, continúan surgiendo estudios sobre modificaciones y aportaciones metodológicas, como el uso del agrupamiento de K-means para validación cruzada contra la metodología común (Saatchi *et al.*, 2011). La evaluación genómica en un solo paso (ssGBLUP) es un análisis estadístico novedoso en SG que permite utilizar información genómica y de pedigrí, con precisión superior en comparación con otros métodos de varios pasos (Wang *et al.*, 2014). El análisis ssGBLUP requiere la construcción de varias matrices: la matriz A (de relaciones genéticas aditivas),

validation against the common methodology (Saatchi *et al.*, 2011). Single-step genomic BLUP evaluation (ssGBLUP) is a novel statistical analysis in GS that allows the use of genomic and pedigree information, with superior accuracy compared to other multi-step methods (Wang *et al.*, 2014). The ssGBLUP analysis requires the construction of several matrix arrays: matrix A (of additive genetic relationships), matrix G (of genomic relationships) and matrix H (of relationships between genotyped and non-genotyped animals) (Pértille *et al.*, 2016).

Regardless of the selected statistical methodologies for the analysis of the information, it is required the provision of high-quality information. Anderson *et al.* (2010) detail the methodology to perform a good genotype quality control (GQC), one of the steps is the elimination of markers with very low Minor Allele Frequency (MAF). MAF is defined as all those alleles, in the DNA sequence, whose frequency ranges from 0.01 to 0.2 (Chouraki & Seshadri, 2014). MAF has also been related to the prevalence of rare genetic diseases (Kido *et al.*, 2018). These authors reported that 50 % of MAF is associated with complex diseases, such as Alzheimer's disease, diabetes, cancer, and schizophrenia.

In the field of genetic edition, the elimination of MAF is intended to increase the frequency of alleles that have an important effect on the traits of interest. To ensure genotype quality in genetic value predictions, markers with MAF less than 5 % or those up to 1 % are removed. For example, Minozzi *et al.* (2012) worked on the GQCs of bovine genotypes, eliminating markers with MAF less than 1 %. Anderson *et al.* (2010) indicated that failure to remove low-frequency alleles results in the use of false information, for two reasons 1) the associations observed in these SNPs are small because they are addressed by the genotypes of few individuals; and 2) they may come from genotyping errors in markers that are actually monomorphic in the population.

For growth traits in cattle, the effect of the MAF threshold on the accuracy of predictions in small populations is not completely clear, although it is understood that increasing genotype quality will also increase the accuracy of GV predictions, even when using high MAF values implies the elimination of several SNP marker information. Therefore, the present research aimed to compare the effect of different MAF thresholds in GQC editing using SNPs for single-stage genomic prediction of growth traits in Swiss-European cattle.

la matriz G (de relaciones genómicas) y la matriz H (de relaciones entre animales genotipados y no genotipados) (Pértille *et al.*, 2016).

Independientemente del uso de metodologías estadísticas para el análisis de la información, se requiere que ésta sea de buena calidad. Anderson *et al.* (2010) detallan la metodología para realizar un buen control de calidad del genotipo (GQC), una de las etapas es la eliminación de los marcadores con frecuencia del alelo menor o *Minor Allele Frequency* (MAF) muy baja. La MAF se define como todos aquellos alelos, en la secuencia de ADN, cuya frecuencia varía de 0.01 a 0.2 (Chouraki & Seshadri, 2014). La MAF también se ha relacionado con la prevalencia de enfermedades genéticas raras (Kido *et al.*, 2018). Estos autores reportaron que 50 % de MAF se asocia con enfermedades complejas, como Alzheimer, diabetes, cáncer y esquizofrenia.

En la edición de la información genética, al eliminar los alelos de menor frecuencia se pretende aumentar la frecuencia de los alelos que tienen un efecto importante en las características en estudio. Para asegurar la calidad del genotipo en las predicciones de valores genéticos se eliminan los marcadores con MAF menor que 5 % o hasta 1 %. Por ejemplo, Minozzi *et al.* (2012) trabajaron en el GQC de genotipos bovinos, eliminando marcadores con MAF menor que 1 %. Anderson *et al.* (2010) indican que el no eliminar los alelos de baja frecuencia provoca el uso de información falsa, por dos razones 1) las asociaciones observadas en estos SNPs son pequeñas porque son impulsadas por los genotipos de pocos individuos; y 2) pueden provenir de errores de genotipado en los marcadores que en realidad son monomórficos en la población.

Para características de crecimiento en bovinos no existe claridad completa del efecto del umbral de MAF en la precisión de las predicciones en poblaciones pequeñas, aunque se entiende que al aumentar la calidad del genotipo se incrementará también la precisión en las predicciones de los VG, aun cuando utilizar valores altos de MAF implique eliminar gran parte de la información de marcadores SNP. Por lo anterior, el objetivo del presente estudio fue comparar el efecto de diferentes umbrales MAF en la edición de GQC utilizando SNPs, para la predicción genómica en una sola etapa de características de crecimiento en bovinos Suizo Europeo.

Material y Métodos

Origen de la información

La información fenotípica provino de la base de

Material and Methods

Phenotypic information source

Phenotypic information was obtained from the database of the Asociación Mexicana de Criadores de Ganado Suizo de Registro (Mexican Association of Breeders of Swiss Cattle of Registration). The pedigree records, identification and birth (BW) and weaning weights (WW) of the animals were taken. The pedigree consisted of 184,788 animals born between 1901 and 2016.

For the genomic information, hair samples were collected from 300 animals, 236 females and 64 males, randomly selected from cattle ranches located in the states of Colima, Jalisco and Veracruz, Mexico. The samples were sent to GeneSeek (Lincoln, NE, USA) for genotyping. The chip used was Genomic Profile Bovine LD, of 30,000 and 50,000 SNP markers for 150 animals each.

Phenotypes

For BW, records of individuals with a BW value outside the mean \pm three standard deviations interval, without herd or maternal age information, were eliminated. Contemporary groups were defined by combining the effects of sex, herd, year and time of birth. Season of birth were defined considering the Julian day: 80-171, spring; 172-264, summer; 265-354, fall; 355-366, winter; and 1-79, winter. Information from animals in contemporary groups with less than two individuals and animals in groups of three or more individuals with zero variance were also deleted. Finally, 28,973 phenotypic records were obtained for further statistical analysis.

For WW, records with inconsistent weaning ages that were three standard deviations above or below the mean, with no information on management, herd, or age of dam, were discarded. Contemporary groups were defined by combining the effects of management, sex, herd, year, and time of birth. Management for WW was defined in three ways: 1) calf feeding with dam's milk; 2) dam's milk plus balanced feed; and 3) dam's milk and a wet nurse plus balanced feed. The season of birth was defined in the same way as for BW, as well as the criterion for elimination within contemporary groups. Finally, 18,994 phenotypic records were obtained for subsequent analysis.

The phenotypes selected for analysis with the ssGBLUP method were from the genotyped animals and their ancestors five generations back. The amount of data on which these analyses were performed is shown in Table 1.

datos de la Asociación Mexicana de Criadores de Ganado Suizo de Registro. Se tomaron los registros genealógicos, identificación y los pesos al nacimiento (BW) y al destete (WW) de los animales. El pedigrí lo conformaron 184,788 animales nacidos entre 1901 y 2016.

Para la información genómica se colectaron muestras de pelo de 300 animales, 236 hembras y 64 machos, seleccionados aleatoriamente de ranchos ubicados en los estados de Colima, Jalisco y Veracruz, México. Posteriormente las muestras se enviaron a genotipar a GeneSeek (Lincoln, NE, USA). El chip usado fue Genomic Profile Bovine LD, de 30,000 y 50,000 marcadores SNP para 150 animales cada uno.

Fenotipos

Para BW, se eliminaron los registros de individuos con BW fuera del intervalo promedio \pm tres desviaciones estándar, sin información de hato o edad de la madre. Los grupos contemporáneos se definieron combinando los efectos de sexo, hato, año y época de nacimiento. Las épocas de nacimiento se definieron considerando el día juliano: del día 80-171, primavera; 172-264, verano; 265-354, otoño; 355-366, invierno; y 1-79, invierno. La información de animales en grupos contemporáneos con menos de dos individuos y animales en grupos de tres o más individuos con varianza cero también se eliminó. Finalmente se obtuvieron 28,973 registros fenotípicos para análisis posteriores.

Para WW se eliminaron los registros con edades al destete inconsistentes, que estuvieron tres desviaciones estándar por arriba o por debajo de la media, sin información de manejo, hato o edad de la madre. Los grupos contemporáneos se definieron combinando los efectos de manejo, sexo, hato, año y época de nacimiento. El manejo para WW se definió de tres maneras 1) alimentación de la cría con leche de la madre; 2) leche de la madre más alimento balanceado; y 3) leche de la madre y de una nodriza más alimento balanceado. Las épocas de nacimiento se definieron igual que para BW, así como el criterio de eliminación dentro de grupos contemporáneos. Finalmente se obtuvieron 18,994 registros fenotípicos para análisis subsecuentes.

Los fenotipos seleccionados para el análisis con el método ssGBLUP fueron de los animales genotipados y de sus ancestros cinco generaciones atrás. La cantidad de datos con que se realizaron estos análisis se muestran en el Tabla 1.

Table 1.
Number of genotyped and non-genotyped animals by birth weight (BW) and weaning weight (WW) analyzed in this study.

Tabla 1.
Número de animales genotipados y no genotipados utilizados en el estudio para los pesos al nacimiento (BW) y al destete (WW).

Trait	Animals	Genotyped	Non-genotyped
BW	325	226	99
WW	255	206	49

Genotypes

Genotype information was edited as follows: *genomic information*: SNPs in common were taken from the 30K and 50K chips, yielding 12,835 SNPs from the 300 animals; *recoding*: additive effects were recoded as AA=0, AB=1 and BB=2; *imputation*: missing genotypes were imputed within chip size using the marginal distribution samples of marker genotypes; and *genotype quality*: monomorphic SNPs and with different MAF thresholds were removed. Table 2 shows the numbers of SNP markers available once the MAF thresholds studied were applied.

Genotipos

La información de los genotipos se editó de la siguiente manera: *información genómica*: se tomaron los SNPs en común del chip de 30K y del de 50K, obteniéndose 12,835 SNPs de los 300 animales; *recodificación*: los efectos aditivos se recodificaron como AA=0, AB=1 y BB=2; *imputación*: los genotipos faltantes se imputaron dentro de tamaño de chip utilizando las muestras de distribución marginal de los genotipos de los marcadores; y *calidad del genotipo*: se eliminaron los SNPs monomórficos y con diferentes umbrales de MAF. En la Tabla 2 se observan los números de marcadores SNP disponibles una vez que se aplicaron los umbrales MAF estudiados.

Table 2.
Minor allele frequency (MAF) and number of available single nucleotide polymorphisms SNP markers for birth and weaning weights analysis.

Tabla 2.
Frecuencia de alelo menor (MAF) y número de marcadores SNP disponibles para el análisis de los pesos al nacimiento y al destete.

MAF	SNP Markers
0	12,557
0.02	11,496
0.04	8,629
0.05	6,648

Calculation of matrix H

The matrix of relationships between genotyped and non-genotyped animals (**H**) was obtained in a similar way as described by Christensen & Lund (2010) and Legarra *et al.* (2009). Matrix **H**, which includes pedigree and genomic information, is defined as:

Cálculo de la matriz H

La matriz de relaciones entre animales genotipados y no genotipados (**H**) se obtuvo de manera similar a como lo descrito por Christensen & Lund (2010) y Legarra *et al.* (2009). La matriz **H**, que incluye información de pedigrí y genómica, se define como:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{nn} + \mathbf{A}'_{gn}\mathbf{A}_{gg}^{-1}(\mathbf{G}_a - \mathbf{A}_{gg})\mathbf{A}_{gg}^{-1}\mathbf{A}_{gn} & \mathbf{A}'_{gn}\mathbf{A}_{gg}^{-1}\mathbf{G}_a \\ \mathbf{G}_a\mathbf{A}_{gg}^{-1}\mathbf{A}_{gn} & \mathbf{G}_a \end{bmatrix}$$

where \mathbf{A}_{nn} , \mathbf{A}_{gn} and \mathbf{A}_{gg} are submatrices of \mathbf{A} containing the relationships between non-genotyped animals, between genotyped with non-genotyped and between genotyped, respectively.

\mathbf{G}_a is a fitted relationship matrix obtained from the matrix \mathbf{G} described above:

$$\mathbf{G}_a = \beta \mathbf{G} + \alpha$$

$\mathbf{G}_a = \beta \mathbf{G} + \alpha$ and are obtained by solving the system of equations:

In the first equation the average of the diagonal

$$\begin{aligned} \text{Avg}(\mathbf{G})\beta + \alpha &= \text{Avg}(\mathbf{A}_{gg}) \\ \text{Avg}(\text{diag}(\mathbf{G}))\beta + \alpha &= \text{Avg}(\text{diag}(\mathbf{A}_{gg})) \end{aligned}$$

elements of \mathbf{G} is equal to the average of the diagonal elements of \mathbf{A}_{gg} . In the second equation the average of all elements of \mathbf{G} is equal to the average of all elements of \mathbf{A}_{gg} . For more details on the calculation of \mathbf{H} consult (Christensen *et al.*, 2012).

Statistical analysis

The BW and WW growth characteristics were analyzed under the one-step genomic analysis approach, modification of the Genomic Best Linear Unbiased Prediction method (GBLUP) proposed by (VanRaden *et al.*, 2009), the most important change being the permutation of matrix \mathbf{G} , by matrix \mathbf{H} . In matrix notation the model can be written as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{c} + \mathbf{Z}_2\mathbf{u} + \mathbf{e}$$

where \mathbf{y} is the vector of phenotypes; \mathbf{Z}_1 and \mathbf{Z}_2 are incidence matrices relating phenotypic measurements to fixed effects and random effects and \mathbf{e} is the vector of residual error random effects. The genetic, contemporary group and residual additive variances were assumed:

$$\text{Var}(\mathbf{u}) = \mathbf{H}\sigma_a^2 \quad \text{Var}(\mathbf{c}) = \mathbf{I}\sigma_g^2 \quad \text{and} \quad \text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$$

respectively.

The model for BW (1) included linear (EM) and quadratic (EM²) maternal age covariate fixed effects; the contemporary group (CG) effect was considered random. The model for WW (2) included the same fixed effects as (1) plus that of the covariate age at weaning (ED).

$$(1) \quad BW = EM + EM^2 + GC + e$$

$$(2) \quad WW = ED + EM + EM^2 + GC + e$$

donde \mathbf{A}_{nn} , \mathbf{A}_{gn} y \mathbf{A}_{gg} son submatrices de \mathbf{A} que contienen las relaciones entre animales no genotipados, entre genotipados con no genotipados y entre genotipados, respectivamente.

\mathbf{G}_a es una matriz de relación ajustada obtenida de la matriz \mathbf{G} descrita anteriormente:

$\mathbf{G}_a = \beta \mathbf{G} + \alpha$ y se obtienen al resolver el sistema de ecuaciones:

En la primera ecuación el promedio de los elementos de la

diagonal de \mathbf{G} es igual al promedio de los elementos de la diagonal de \mathbf{A}_{gg} . En la segunda ecuación el promedio de todos los elementos de \mathbf{G} es igual al promedio de todos los elementos de \mathbf{A}_{gg} . Para más detalles sobre el cálculo de \mathbf{H} consultar (Christensen *et al.*, 2012).

Análisis estadístico

Las características de crecimiento BW y WW se analizaron bajo el enfoque de análisis genómico en un solo paso, modificación del método de mejor predicción genómica insesgada genómica (GBLUP) propuesto por (VanRaden *et al.*, 2009), siendo el cambio más importante la permutación de la matriz \mathbf{G} , por la matriz \mathbf{H} . En notación matricial el modelo puede escribirse como:

donde \mathbf{y} es el vector de fenotipos; \mathbf{Z}_1 y \mathbf{Z}_2 son matrices de incidencia que relacionan las mediciones fenotípicas con los efectos fijos y los efectos aleatorios y \mathbf{e} es el vector de efectos aleatorios de error residual. Las varianzas aditivas genéticas, de grupo contemporáneo y residual se supusieron:

$$\text{Var}(\mathbf{u}) = \mathbf{H}\sigma_a^2, \quad \text{Var}(\mathbf{c}) = \mathbf{I}\sigma_g^2 \quad \text{y} \quad \text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$$

respectivamente.

El modelo para BW (1) incluyó los efectos fijos de la covariable edad de la madre lineal (EM) y cuadrático (EM²); el efecto de grupo contemporáneo (GC) fue considerado aleatorio. El modelo para WW (2) incluyó los mismos efectos fijos que (1) más el de la covariable edad al destete (ED).

$$(1) \quad BW = EM + EM^2 + GC + e$$

$$(2) \quad WW = ED + EM + EM^2 + GC + e$$

Cross-validation

The predictive ability of the models used was tested by cross-validation. The population was divided into five groups, one of which served as training and the rest as test groups. The cross-validation was repeated four times, with each of the test data groups, and finally the arithmetic mean of the results of each iteration was obtained, obtaining a single result for each analysis.

Criteria for comparison of analyses

MAF Effect on ranking of the animals

To determine the effect of MAF on the degree of similarity between animal ranking according to genomic estimated breeding values (GEBVs) for BW and WW, Spearman and Pearson correlation estimators were obtained between the GEBVs obtained with MAF = 0.05 and with the corresponding GEBVs obtained with MAF = 0, 0.02 and 0.04.

Predictive ability of the models

To evaluate the effect of MAF on the predictive ability of the models, a cross-validation was performed. The data were divided into training and validation sets, using the validation set to evaluate the predictive ability of the trained model. Twenty percent of the phenotypic values of the BW and WW growth traits were randomly removed. Subsequently for each model with MAFs of 0, 0.02, 0.04 and 0.05, the GEBVs for those traits were obtained and the average of four replicates of the correlation coefficients of the deleted phenotypes with the predicted ones was obtained.

Simple linear regression coefficient

Simple linear regression coefficient estimators were used to determine the effect of MAF on changes in magnitude of GEBVs, using GEBVs with MAF 0.05 as independent variables and genomic estimated breeding values predicted with MAF 0, 0.02, and 0.04 as dependent variables. The null hypothesis tested was that the regression coefficient estimator was equal to zero.

R software (RStudio Team, 2019) was used for all statistical analyses.

Results and discussion

Ranking of the animals

Spearman's and Pearson's correlation coefficients

Validación cruzada

La habilidad predictiva de los modelos utilizados se probó mediante validación cruzada. La población se dividió en cinco grupos, uno de los cuales sirvió como de entrenamiento y el resto como de prueba. La validación cruzada se repitió cuatro ocasiones, con cada uno de los grupos de datos de prueba y finalmente se obtuvo la media aritmética de los resultados de cada iteración obteniendo un único resultado para cada análisis.

Criterios de comparación de los análisis

Efecto de MAF en la jerarquización de animales

Para determinar el efecto de MAF sobre el grado de similitud entre la jerarquización de animales de acuerdo con los valores genómicos predichos (GEBVs) para BW y WW, se obtuvieron los estimadores de correlación Spearman y Pearson entre los GEBVs obtenidos con MAF = 0.05 y con los correspondientes GEBVs obtenidos con MAF = 0, 0.02 y 0.04.

Habilidad de predicción de los modelos

Para evaluar el efecto de MAF en la habilidad predictiva de los modelos se realizó una validación cruzada. Se dividió la información en conjuntos de entrenamiento y validación, utilizando el conjunto de validación para evaluar la capacidad predictiva del modelo entrenado. Se eliminó aleatoriamente 20 % de los valores fenotípicos de las características de crecimiento BW y WW. Posteriormente para cada modelo con MAF de 0, 0.02, 0.04 y 0.05 se obtuvieron los GEBVs para dichas características y se obtuvo el promedio de cuatro repeticiones de los coeficientes de correlación de los fenotipos eliminados con los predichos.

Coefficiente de regresión lineal simple

Los estimadores de los coeficientes de regresión lineal simple se utilizaron para determinar el efecto de MAF en los cambios en magnitud de los GEBVs, utilizando como variables independientes los GEBVs con MAF 0.05 y como variables dependientes los valores genómicos predichos con MAF 0, 0.02 y 0.04. La hipótesis nula probada fue que el estimador del coeficiente de regresión fue igual que cero.

Para todos los análisis estadísticos se utilizó el programa R (RStudio Team, 2019).

Resultados y discusión

Jerarquización de los animales

Los coeficientes de correlación de Spearman y Pearson

between GEBVs with the analyses considering different MAF thresholds are shown in Table 3. The estimators in both types of correlations were greater than 0.999 ($p \leq 2 \times 10^{-16}$), there were minor changes in the animal ranking according to the GEBVs obtained with different thresholds tested. This indicates that the effect of MAF was not significant in the ranking of GEBVs.

These results suggest that for purposes of selecting animals as parents, when evaluations are done using SNP information, any MAF threshold between 0.0 and 0.05 can be used in quality control during genotype editing, although 0.05 is the standard in genomic evaluations performed with animals according to (VanRaden *et al.*, 2009). The accuracies obtained by

entre los GEBVs con los análisis considerando diferentes umbrales de MAF se muestran en la Tabla 3. Los estimadores en ambos tipos de correlaciones fueron mayores que 0.999 ($p \leq 2 \times 10^{-16}$), hubo pocos cambios en la jerarquización de los animales de acuerdo con los GEBVs obtenidos con los diferentes umbrales probados. Esto indica que el efecto de MAF no fue importante en la jerarquización de los GEBVs.

Estos resultados sugieren que con fines de selección de animales como progenitores, cuando las evaluaciones se hacen utilizando información de SNPs, se puede utilizar cualquier umbral MAF entre 0.0 y 0.05 en el control de calidad durante la edición de genotipos, aunque 0.05 es el estándar en las evaluaciones genómicas realizadas con animales de acuerdo con (VanRaden *et al.*, 2009).

Table 3.
Estimation of Pearson and Spearman correlation coefficients, in parenthesis significance level for the hypothesis test of estimate equal to zero between genomic values predicted with minor allele frequency (MAF) = 0.05 versus MAF = 0, 0.02, and 0.04 for birth (BW) and weaning (WW) weights in Braunvieh cattle.

Tabla 3.
Estimadores de los coeficientes de correlación Pearson y Spearman, en paréntesis nivel de significancia para la prueba de hipótesis del estimador igual que cero, entre valores genómicos predichos con frecuencia del alelo menor (MAF) = 0.05 contra MAF = 0, 0.02 y 0.04 para pesos al nacimiento (BW) y al destete (WW) en bovinos Suizo Europeo.

MAF	BW		WW	
	Pearson	Spearman	Pearson	Spearman
0	0.99993	0.99980	0.99976	0.99960
	(<2×10 ⁻¹⁶)	(<2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)
0.02	0.99990	0.99976	0.99974	0.99957
	(<2×10 ⁻¹⁶)	(<2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)
0.04	0.99991	0.99978	0.99983	0.99968
	(<2×10 ⁻¹⁶)	(<2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)	(<2.2×10 ⁻¹⁶)

MAF = threshold for minor allele frequency; BW = birth weight; WW = weaning weight.

MAF = umbral de frecuencia de alelo menor; BW = peso al nacimiento; WW = peso al destete.

correlation estimators were slightly higher for BW than for WW, this may be due to the longer time of environmental influence on animals for WW and that may affect the predictions.

In genotyping quality control in Australian Holstein-Friesian cattle, using the Illumina Bovine SNP50TM chip and 798 animals, MAF <2.5 % was eliminated (Hayes *et al.*, 2009). These authors obtained live weight accuracies for GEBVs of 50 to 67 %, similar to

Las precisiones obtenidas mediante los estimadores de correlación fueron ligeramente mayores para BW que para WW, esto puede deberse al mayor tiempo de influencia del ambiente sobre los animales para WW y que puede afectar las predicciones.

En el control de calidad del genotipo en bovinos Holstein-Friesian australianos, usando el chip Illumina Bovine SNP50TM y 798 animales, se eliminó MAF <2.5 % (Hayes *et al.*, 2009). Estos autores obtuvieron para peso vivo exactitudes de los GEBVs de 50 a 67 %, similares a las

those obtained in the present research. Wiggans *et al.* (2009) eliminated records using MAF less than 2 % in genotypes of 941 Jersey cattle and 344 Braunvieh bulls, the two populations of animals used in these studies cannot be considered large according to VanRaden *et al.* (2009). These authors suggest using MAF no less than 5 %, although in large populations it can be decreased without affecting accuracy.

Wiggans *et al.* (2009) argue that the number of animals available per breed in their study meant that their results were considered preliminary, as the sample size did not allow definitive results to be obtained. This may explain the results obtained in the present investigation, as the number of animals used is not considered a large population compared to the population sizes used in other studies (VanRaden *et al.*, 2009).

Predictive ability of the model with different MAF thresholds

Table 4 shows that setting a higher MAF in the quality control in genotype editing improves the prediction accuracy of GEBVs, although the difference in accuracy, for BW, between using MAF = 0.0 versus MAF = 0.05 is very low, 0.0027, and even lower for WW where the difference was 0.0007. This means that, in the GQC of a sample of 300 animals with 12,835 SNPs, the MAF

obtenidas en el presente estudio. Por su parte, Wiggans *et al.* (2009) eliminaron registros utilizando MAF menores que 2 % en genotipos de 941 bovinos Jersey y 344 toros Suizo Europeo, las dos poblaciones de animales utilizadas en estos estudios no pueden considerarse grandes de acuerdo con VanRaden *et al.* (2009). Estos autores sugieren usar MAF no menor que 5 %, aunque en poblaciones grandes se puede disminuir sin afectar la precisión.

Wiggans *et al.* (2009) argumentan que el número de animales disponibles por raza en su estudio hizo que sus resultados se consideraran preliminares, ya que el tamaño de muestra no permitió obtener resultados definitivos. Lo anterior puede explicar los resultados obtenidos en la presente investigación, pues la cantidad de animales utilizados no se considera una población grande, en comparación con los tamaños poblacionales utilizados en otros estudios (VanRaden *et al.*, 2009).

Habilidad de predicción del modelo con diferente umbral de MAF

La Tabla 4 muestra que establecer un MAF mayor en el control de calidad en la edición de los genotipos, mejora la precisión de predicción de los GEBVs, aunque la diferencia en precisión, para BW, entre utilizar MAF = 0.0 contra MAF = 0.05 es muy baja, 0.0027, y es más baja para WW donde la diferencia fue 0.0007. Esto significa que, en el GQC de una muestra de 300 animales con 12,835 SNPs,

Table 4.
Averages for Pearson correlation coefficient between genomic values predicted analyzing different thresholds for minor allele frequency (MAF) and the actual phenotype values for weights at birth (BW) and at weaning (WW) of four-fold cross validation.

Tabla 4.
Promedios de los estimadores de correlación Pearson entre los valores genómicos predichos mediante análisis con diferentes umbrales de frecuencia del alelo menor (MAF) y los valores fenotípicos reales para los pesos al nacimiento (BW) y al destete de cuatro repeticiones de validación cruzada (WW).

MAF	BW	WW
0	0.7130	0.6444
0.02	0.7127	0.6447
0.04	0.7136	0.6451
0.05	0.7136	0.6451

MAF = threshold for minor allele frequency; BW = birth weight; WW = weaning weight.
MAF = umbral de frecuencia de alelo menor; BW = peso al nacimiento; WW = peso al destete.

threshold to be used of up to 0.05 is not significant in the prediction of GEBVs for BW and WW.

Linck & Battey (2019) conducted a study with simulated SNP information, using MAF thresholds between 0.017 and 0.25; these authors reported that using a MAF greater than 0.03 implies a change in the structure of the data distribution due to the drop in the total size of the matrix after filtering and generates an increase in the accuracy of the predictors. This may explain why, in the present study, using MAF = 0.05 for BW implies slightly better predictions.

The predictive ability of the models for BW and WW for each of the thresholds studied does not represent significant changes. The above can be explained by the small number of SNPs removed given the information used compared to the information used in other studies such as that performed by (Zhu *et al.*, 2017) with 82,594 markers by (De la Cruz & Raska, 2014), 40,000 in the study by (Linck & Battey, 2019), and 600,000 in the bird study by (Abdollahi-Arpanahi *et al.*, 2014). These studies reported a significant effect of MAF on genomic prediction of the studied traits. An additional aspect is the low number of animals used in the present study, which does not allow conclusive results. A broader analysis should be carried out using a larger animal population.

Estimator of regression coefficients

Table 5 shows that the regression coefficients are similar to each other; however, the coefficient for WW decreases slightly as the MAF for quality control in genotype editing increases. As the estimators of the regression coefficients are different from unity, the difference in the prediction of GEBVs with the analysis alternatives in comparison is greater (Mäntysaari *et al.*, 2010), all the estimators of the regression coefficients are close to unity consequently the differences between the effects of the MAF thresholds tested are minimal.

The results of the present study are similar to those reported by (Edriss *et al.*, 2012), who using information from 5,500 Holstein and Jersey animals with 54,000 SNPs, estimated for fertility, milk protein and mastitis traits, the correlations between observed and predicted values using MAF thresholds: 0.075, 0.275, 0.3, 0.325 and 0.35, their results showed minimal differences, for example, in Holstein the correlations for fertility had

el umbral de MAF a utilizar de hasta 0.05 no es importante en la predicción de GEBVs para BW y WW.

Linck & Battey (2019) realizaron un estudio con información de SNPs simulada, utilizando umbrales de MAF entre 0.017 y 0.25; estos autores reportaron que utilizar un MAF mayor que 0.03 implica un cambio en la estructura de la distribución de los datos por la caída en el tamaño total de la matriz después del filtrado y genera un aumento en la precisión de los predictores. Lo anterior puede explicar que en el presente estudio utilizar MAF = 0.05 para BW, implique obtener ligeramente mejores predicciones.

La habilidad de predicción de los modelos para BW y WW para cada uno de los umbrales estudiados no representa cambios significativos. Lo anterior puede explicarse por la poca cantidad de SNPs eliminados dada la información utilizada comparada con la información que se usó en otros estudios como el realizado por (Zhu *et al.*, 2017) con 82,594 marcadores por (De la Cruz & Raska, 2014), 40,000 en el estudio de (Linck & Battey, 2019), y 600,000 en el estudio con aves de (Abdollahi-Arpanahi *et al.*, 2014). Estos autores reportaron un efecto importante del MAF en la predicción genómica de las características estudiadas. Un aspecto adicional es el bajo número de animales que se utilizó en el presente estudio, lo que no permite obtener resultados concluyentes. Se recomienda realizar el estudio con un número mayor de animales.

Estimador de los coeficientes de regresión

En la Tabla 5 se observa que los coeficientes de regresión son similares entre sí; sin embargo, el coeficiente para WW se reduce ligeramente conforme incrementa el MAF para el control de calidad en la edición del genotipo. Conforme los estimadores de los coeficientes de regresión sean diferentes de la unidad, la diferencia en la predicción de los GEBVs con las alternativas de análisis en comparación es mayor (Mäntysaari *et al.*, 2010), todos los estimadores de los coeficientes de regresión se encuentran cerca de la unidad en consecuencia las diferencias entre los efectos de los umbrales de MAF probados son mínimas.

Los resultados del presente estudio son similares a los reportados por (Edriss *et al.*, 2012), quienes utilizando información de 5,500 animales Holstein y Jersey con 54,000 SNPs, estimaron para las características de fertilidad, proteína en leche y mastitis, las correlaciones entre valores reales y predichos usando los umbrales de MAF: 0.075, 0.275, 0.3, 0.325 y 0.35, sus resultados mostraron mínimas diferencias, por ejemplo, en Holstein las correlaciones para fertilidad tuvieron una diferencia máxima de 0.014;

Table 5.

Estimation of regression coefficients (\pm SE), independent variables were the genomic values predicted with minor allele frequency (MAF) = 0.05 and dependent variables were the genomic values predicted with MAF = 0, 0.02, and 0.04; in parenthesis significance levels for hypothesis test of estimate equal to zero for birth (BW) and weaning (WW) weights in Braunvieh cattle.

Tabla 5.

Estimadores de los coeficientes de regresión (\pm SE), las variables independientes fueron los valores genómicos predichos con frecuencia del alelo menor (MAF) = 0.05 y como las dependientes los valores genómicos predichos con MAF de 0, 0.02 y 0.04; entre paréntesis nivel de significancia para la prueba de hipótesis del estimador igual que cero, para los pesos al nacimiento (BW) y al destete (WW) en bovinos Suizo Europeo.

MAF	BW	WW
0	0.99756 \pm 0.00072 ($<2 \times 10^{-16}$)	0.99323 \pm 0.00136 ($<2 \times 10^{-16}$)
0.02	0.99605 \pm 0.00076 ($<2 \times 10^{-16}$)	0.99255 \pm 0.00141 ($<2 \times 10^{-16}$)
0.04	0.99578 \pm 0.00072 ($<2 \times 10^{-16}$)	0.99400 \pm 0.00114 ($<2 \times 10^{-16}$)

MAF = threshold for minor allele frequency; BW = birth weight; WW = weaning weight.

MAF = umbral de frecuencia de alelo menor; BW = peso al nacimiento; WW = peso al destete.

a maximum difference of 0.014; concluding that the results of their study suggest that for a small reference population, SNP markers with low MAF do not impair genomic prediction, a statement that also applies to the findings of the present study, but for growth traits.

As reported by VanRaden *et al.* (2009) and Wiggans *et al.* (2009), in relation to the number of animals in the population and the number of markers, the results obtained in this study should be taken with caution, and as mentioned by Hayes *et al.* (2009), they should be considered preliminary and susceptible to changes that may occur in the future when repeating the study with a larger number of animals and markers, since the effect of MAF is due to the drop in the total matrix size after filtering in the GQC, a drop that may not be significant if the amount of information is limited (Linck & Battey, 2019).

Conclusion

The minor allele frequency threshold to use in genomic evaluation could be in the range of 0 to 0.05, without significantly altering the prediction of genomic values, as it does not have a significant effect on the

concluyendo que los resultados de su estudio sugieren que para una población de referencia pequeña, los marcadores SNP con MAF bajo no dañan la predicción genómica, afirmación que también aplica a lo encontrado en el presente estudio, pero para características de crecimiento.

De acuerdo con lo reportado por VanRaden *et al.* (2009) y Wiggans *et al.* (2009), en relación con el número de animales que conforman la población y el número de marcadores, los resultados utilizados en este estudio deberán tomarse con precaución, y como mencionaron Hayes *et al.* (2009), deberán considerarse preliminares y susceptibles a cambios que en un futuro pudieran presentarse al repetir el estudio con un número mayor de animales y marcadores, ya que el efecto de MAF se debe a la caída en el tamaño total de la matriz después del filtrado en la GQC, caída que puede no ser significativa si la cantidad de información es limitada (Linck & Battey, 2019).

Conclusión

El umbral de frecuencia de alelo menor a utilizar en la evaluación genómica puede encontrarse en el rango de 0 a 0.05, sin alterar de forma importante la predicción de valores genómicos, pues no tiene un

ranking of predicted genomic values and the predictive ability of the model. In this sense, it is concluded that 0.05 may be a recommendable threshold to use in the evaluation of small populations and with a limited number of markers.

efecto significativo sobre la jerarquización de los valores genómicos predichos y la habilidad predictiva del modelo. Se reconoce que 0.05 puede ser un umbral recomendable para utilizar en evaluación de poblaciones pequeñas y con un número de marcadores limitado.

References

- Abdollahi-Arpanahi, R., Nejati-Javaremi, A., Pakdel, A., Moradi-Shahrbabak, M., Morota, G., Valente, B. D., Kranis, A., Rosa, G. J. M. and Gianola, D. (2014). Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. *Journal of Animal Breeding and Genetics*, 131(2): 123-133. <https://doi.org/10.1111/jbg.12075>
- Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P. and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*, 5, 1564-1573. <https://doi.org/10.1038/nprot.2010.116>
- Christensen, O. F. & Lund, M. S. (2010). Genomic relationship matrix when some animals are not genotyped. *Genetics Selection Evolution*, 42(2): 1-8. <https://doi.org/10.1186/1297-9686-42-2>
- Christensen, O. F., Madsen, P., Nielsen, B., Ostensen, T. and Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal* 6(10): 1565-1571. <https://doi.org/10.1017/S1751731112000742>
- Chouraki, V. & Seshadri, S. (2014). Chapter five - Genetics of Alzheimer's disease. *Advances in Genetics*, 87, 246-278. <https://doi.org/10.1016/B978-0-12-800149-3.00005-6>
- De la Cruz, O. & Raska, P. (2014). Population structure at different minor allele frequency levels. *BMC Proc*, 8, S55. <https://doi.org/10.1186/1753-6561-8-s1-s55>
- De los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D. and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2): 327-345. <https://doi.org/10.1534/genetics.112.143313>
- Edriss, V., Guldbrandtsen, B., Lund, M. S. and Su, G. (2012). Effect of marker-data editing on the accuracy of genomic prediction. *Journal of Animal Breeding and Genetics*, 130(2): 128-135. <https://doi.org/10.1111/j.1439-0388.2012.01015.x>
- Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E. and Fernando, R. (2009). Additive Genetic Variability and the Bayesian Alphabet. *Genetics*, 183(1): 347-363. <https://doi.org/10.1534/genetics.109.103952>
- Gianola, D., Fernando, R. L. and Stella, A. (2006). Genomic-Assisted Prediction of Genetic Value with Semiparametric Procedures. *Genetics*, 173(3): 1761-1776. <https://doi.org/10.1534/genetics.105.049510>
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2): 433-443. <https://doi.org/10.3168/jds.2008-1646>
- Kido, T., Sikora-Wohlfeld, W., Kawashima, M., Kikuchi, S., Kamatani, N., Patwardhan, A., Chen, R., Sirota, M., Kodama, K., Hadley, D. and Butte, A. J. (2018). Are minor alleles more likely to be risk alleles? *BMC Medical Genomics*, 11, 3. <https://doi.org/10.1186/s12920-018-0322-5>
- Legarra, A., Aguilar, I. and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9): 4656-4663. <https://doi.org/10.3168/jds.2009-2061>
- Linck, E. & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3): 639-647. <https://doi.org/10.1111/1755-0998.12995>
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., Singh, R. P., Auriague, E. and De los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *Genes|Genomes|Genetics*, 5(4): 569-582. <https://doi.org/10.1534/g3.114.016097>
- Mäntysaari, E., Liu, Z. and VanRaden, P. (2010). Interbull validation test for genomic evaluations. *INTERBULL BULLETIN* No 41. <https://journal.interbull.org/index.php/ib/article/view/1134>

- Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829. <https://www.genetics.org/content/157/4/1819#:~:text=A%20genome%20of%201000%20cM,marker%20spacing%20of%201%20cM.&text=It%20was%20concluded%20that%20selection,to%20shorten%20the%20generation%20interval>
- Meuwissen, T., Hayes, B. and Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6(1): 6-14. <https://doi.org/10.2527/af.2016-0002>
- Minozzi, G., Williams, J. L., Stella, A., Strozzi, F., Luini, M., Settles, M. L., Taylor, J. F., Whitlock, R. H., Zanella, R. and Neiberghs, H. L. (2012). Meta-analysis of two genome-wide association studies of bovine paratuberculosis. *PLOS ONE*, 7(3): e32578. <https://doi.org/10.1371/journal.pone.0032578>
- Pértile, S. F. N., Silva, F. F., Salvian, M. and Mourão, G. B. (2016). Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. *Pesquisa Agropecuária Brasileira*, 51(10): 1729-1736. <https://doi.org/10.1590/s0100-204x2016001000004>
- RStudio Team. (2019). RStudio: Integrated Development for R. *RStudio, Inc.*, Boston, MA. <https://www.rstudio.com/>
- Saatchi, M., McClure, M. C., McKay, S. D., Rolf, M. M., Kim, J., Decker, J. E., Taxis, T. M., Chapple, R. H., Ramey, H. R., Northcutt, S. L., Bauck, S., Woodward, B., Dekkers, J. C. M., Fernando, R. L., Schnabel, R. D., Garrick, D. J. and Taylor, J. F. (2011). Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution*, 43, 40. <https://doi.org/10.1186/1297-9686-43-40>
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. and Schenkel, F. S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1): 16-24. <https://doi.org/10.3168/jds.2008-1514>
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., Okimoto, R., Wing, T., Hawken, R. and Muir, W. M. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Frontiers in Genetics*, 5, 134. <https://doi.org/10.3389/fgene.2014.00134>
- Wiggans, G. R., Sonstegard, T. S., VanRaden, P. M., Matukumalli, L. K., Schnabel, R. D., Taylor, J. F., Schenkel, F. S. and Van Tassell, C. P. (2009). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *Journal of Dairy Science*, 92, 3431-3436. <https://doi.org/10.3168/jds.2008-1758>
- Zhu, B., Zhang, J., Niu, H., Guan, L., Guo, P., Xu, L., Chen, Y., Zhang, L., Gao, H., Gao, X. and Li, J. (2017). Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. *Journal of Integrative Agriculture*, 16(4): 911-920. [https://doi.org/10.1016/s2095-3119\(16\)61474-0](https://doi.org/10.1016/s2095-3119(16)61474-0)