

# Evaluaciones genéticas en bovinos por medio del uso del mejor predictor lineal insesgado genómico en una etapa

## Genetic evaluations in cattle using the single-step genomic best linear unbiased predictor

Alejandro Amaya Martínez,<sup>1</sup> Rodrigo Martínez Sarmiento,<sup>2</sup> Mario Cerón-Muñoz<sup>3\*</sup>

<sup>1</sup> Docente e investigador, Universidad de Ciencias Aplicadas y Ambientales (UDCA), Facultad de Ciencias Agropecuarias, Bogotá, Colombia. Correo: adamaya@udca.edu.co. Orcid: orcid.org/0000-0002-2783-6567

<sup>2</sup> Director de Investigación y Desarrollo, Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Mosquera, Colombia. Correo: ramartinez@agrosavia.co. Orcid: orcid.org/0000-0003-3550-7103

<sup>3</sup> Docente e investigador, Universidad de Antioquia, Facultad de Ciencias Agrarias, Escuela de Producción Agropecuaria, Grupo GaMMA, Medellín, Colombia. Correo: mario.ceron@udea.edu.co. Orcid: orcid.org/0000-0002-7233-6625

Editor temático: Andrés Castillo Giraldo (Universidad del Valle)

Fecha de recepción: 27/12/2018

Fecha de aprobación: 13/08/2019

Para citar este artículo: Amaya-Martínez, A., Martínez-Sarmiento, R., & Cerón-Muñoz, M. (2020). Evaluaciones genéticas en bovinos por medio del uso del mejor predictor lineal insesgado genómico en una etapa. *Ciencia y Tecnología Agropecuaria*, 21(1), e1548

DOI: [https://doi.org/10.21930/rcta.vol21\\_num1\\_art:1548](https://doi.org/10.21930/rcta.vol21_num1_art:1548)



Esta licencia permite distribuir, remezclar, retocar, y crear a partir de la obra de modo no comercial, siempre y cuando se dé el crédito y se licencien sus nuevas creaciones bajo las mismas condiciones.

\* Autor de correspondencia: Universidad de Antioquia. Carrera 75 No. 65-87, Medellín, Colombia.

## Resumen

Las evaluaciones genéticas convencionales han estado enmarcadas en la estimación de valores genéticos a partir de los sistemas de ecuaciones de modelos mixtos que consideran efectos aleatorios y fijos simultáneamente. En los últimos años, el desarrollo de tecnologías de secuenciación del genoma ha permitido obtener información genómica que puede ser incluida en las evaluaciones genéticas para incrementar las confiabilidades y el progreso genético, así como para disminuir el intervalo generacional. El mejor predictor lineal

insesgado en una etapa es una metodología que incluye información genómica, reemplazando la matriz de parentesco por una matriz que combina el parentesco por pedigrí y el parentesco genómico de una población genotipada, permitiendo la estimación de valores genéticos para animales no genotipados. El objetivo de este artículo de revisión fue la descripción de la metodología y sus recientes avances, así como conocer algunas de las estrategias que podrían ser llevadas a cabo cuando el número de animales genotipados es bajo.

**Palabras clave:** fenotipos, ganadería, genómica, marcadores genéticos, mejoramiento genético

## Abstract

Conventional genetic evaluations have been framed on estimated breeding values from equation systems of mixed models that consider simultaneously random and fixed effects. Recently, the development in genome sequencing technologies has allowed obtaining genomic information to include in genetic evaluations in order to increase the accuracy and genetic progress, and decrease the generation interval. The single-step best linear unbiased predictor is a methodology developed in

the last years and accepts including genomic information replacing the genomic relationship matrix by a matrix that combines relationship by pedigree, and the genomic relationship of a genotyped population, allowing the estimation of breeding values for non-genotyped animals. The aim of this review article was to describe the methodology and its recent progress, as well as to know some of the strategies that could be used when the number of genotyped animals is low.

**Keywords:** animal husbandry, genetic improvement, genetic markers, genomics, phenotypes

## Introducción

Meuwissen, Hayes y Goddard (2001) propusieron el concepto de selección genómica que está basado en la distribución homogénea de miles de marcadores a lo largo del genoma y la estimación de sus efectos sobre características cuantitativas. Los avances tecnológicos y la disponibilidad de chips de polimorfismos de nucleótido simple (SNP, por sus siglas en inglés) de alta densidad han facilitado la implementación de modelos de evaluación genómica (Miształ, Legarra, & Aguilar, 2009). Estos modelos estiman efectos genéticos para los segmentos de cromosomas en una población que cuenta con información fenotípica y genómica, permitiendo conocer los efectos de cada SNP y estimar valores genómicos directos (Pintus et al., 2012).

Paralelamente, el crecimiento acelerado de datos genómicos y los cambios en programas de selección demandan una actualización constante de los sistemas de evaluación genética (VanRaden, 2016). Los modelos de selección genómica han incluido al mejor predictor lineal insesgado (BLUP, por sus siglas en inglés) ampliado, métodos Bayesianos y no paramétricos (Garrick, Dekkers & Fernando, 2014; Howard, Carriquiry & Beavis, 2014). La mayoría de estos modelos explican un mayor porcentaje de la varianza genética aditiva, incrementan las exactitudes de las estimaciones y reducen el intervalo generacional, contribuyendo notoriamente a un mayor progreso genético (Schaeffer, 2006). Por estas razones, la mayoría de estas aproximaciones generalmente han mostrado una superioridad respecto a los métodos de evaluación basados en el uso de la matriz de parentesco por pedigrí y que no incluyen información genómica (Legarra & Ducrocq, 2012).

El uso de estas herramientas genómicas está relacionado con una mayor precisión en las estimaciones y consecuentemente un incremento en la ganancia genética. Las primeras evaluaciones genómicas contemplaron una metodología en múltiples etapas (msGBLUP). Sin embargo, valores genéticos no podían ser estimados para animales sin información genómica en msGBLUP (VanRaden, 2008). Por eso, Miształ et al. (2009) desarrollaron la metodología del mejor predictor lineal insesgado en una etapa

(ssGBLUP) que incorpora información genómica y estima valores genéticos tanto para animales genotipados como para aquellos no genotipados. Por lo tanto, los objetivos de esta revisión fueron los siguientes: 1) describir la metodología ssGBLUP y sus avances; 2) identificar ventajas y limitaciones de la metodología, y 3) conocer la factibilidad de su uso en programas de mejoramiento genético con poblaciones genotipadas pequeñas.

## Método

La consulta de la información bibliográfica se llevó a cabo en las bases de datos de Science Direct, Google Scholar, Scopus, NCBI y Scielo, a partir de 2001, año en que fue publicado el primer artículo de selección genómica referente al uso de SNP en evaluaciones genéticas. Las palabras usadas para la búsqueda fueron *genomic evaluation*, *single step*, *genetic evaluation*, *genomic selection*, *genomic prediction*, *single nucleotide polymorphism* y *genetic modeling*. La revisión no incluyó como fuentes bibliográficas libros ni artículos científicos publicados en un idioma diferente al inglés; tampoco fueron considerados artículos publicados en revistas no especializadas en genética y mejoramiento animal.

## Desarrollo del tema

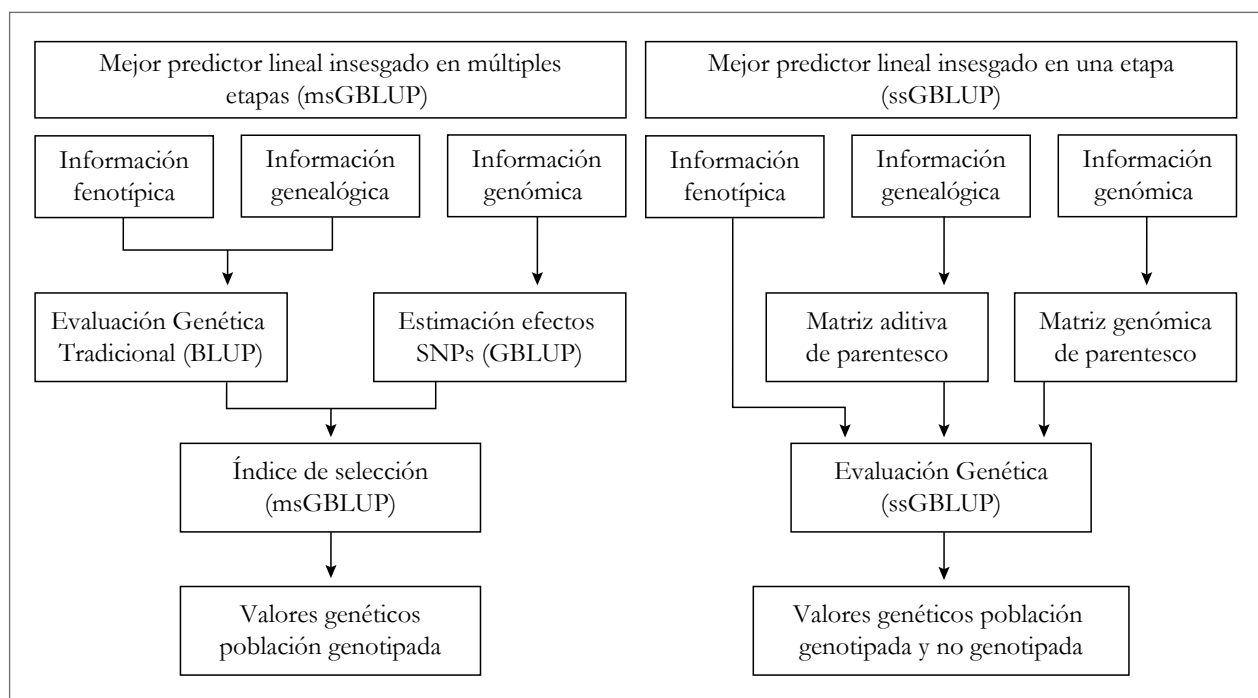
Los programas de mejoramiento genético convencionales han estado basados en la estimación de valores genéticos a partir de información genealógica y fenotípica (Chen et al., 2011). La metodología para estas evaluaciones ha sido el uso de modelos lineales mixtos, los cuales relacionan el desempeño productivo con efectos genéticos y son ajustados por efectos fijos y aleatorios no genéticos (Christensen, Madsen, Nielsen, Ostersen, & Su, 2012). Adicionalmente, la inclusión de la matriz de parentesco basada en el pedigrí permite estimar y ajustar valores genéticos de animales que carecen de información fenotípica, pero que están emparentados con la población evaluada. Sin embargo, el ajuste está basado en probabilidades promedio de segregación mendeliana y podría conducir en algunos casos a la estimación de valores genéticos idénticos para hermanos completos. Esto ignora el coeficiente

de parentesco real y conduce a la subestimación o sobreestimación de valores genéticos para algunos animales.

El esquema anterior ha sido exitoso, pero el progreso genético obtenido es lento si la característica puede ser solamente medida en un sexo, después del sacrificio o tarde en la vida del animal. Para esto, esquemas de selección genómica permitieron incrementar considerablemente la tasa de ganancia genética (Meuwissen et al., 2001). Inicialmente, las evaluaciones genéticas con información genómica se realizaron a partir de paneles con miles de SNP, en el que todos los *loci* están asociados a características cuantitativas y se encuentran en desequilibrio de ligamiento con al menos un SNP (Goddard & Hayes, 2009). Por lo tanto, las decisiones de selección son basadas en los efectos de SNP de cada animal (Moser, Khatkar, Hayes, & Raadsma, 2010). En la actualidad, la selección genómica es ampliamente usada en ganado de carne y leche en países desarrollados como los Estados Unidos. Esto ha sido liderado por asociaciones de criadores y compañías privadas de inseminación artificial, cuyo objetivo es

conducir hacia evaluaciones genéticas que mejoren la habilidad para identificar animales genéticamente superiores (Elzo et al., 2015).

Las primeras evaluaciones genómicas en el mundo usaron msGBLUP (figura 1), que consistió en lo siguiente: 1) la estimación de valores genéticos de manera convencional a través de un modelo animal, 2) estimación de los efectos de los SNP a partir de un grupo de animales con fenotipos y genotipos (población de referencia), y 3) la combinación de la información disponible bajo la teoría del índice de selección para obtener los valores genéticos genómicos (VanRaden, 2008). Sin embargo, la metodología de msGBLUP solo estima valores genéticos para animales genotipados y genera una alta variación de las confiabilidades (Garrick, Taylor, & Fernando, 2009). Aunque esta metodología permitió incrementar la precisión de las estimaciones para la selección de animales jóvenes (König & Swalve, 2009), esta predicción genómica tampoco consideró el efecto de preselección, subestimando valores genéticos para animales jóvenes (Petry & Ducrocq, 2011).



**Figura 1.** Estructura de trabajo del mejor predictor lineal insesgado en múltiples etapas (msGBLUP) y en una sola etapa (ssGBLUP) para la estimación de valores genéticos.

Fuente: Adaptada de Aguilar et al. (2010) y Vanraden (2008)

Para superar dichos obstáculos, ssGBLUP fue desarrollada como una estructura de trabajo unificada que permite realizar estimaciones genómicas en toda la población y más precisas que las obtenidas en msGBLUP (Aguilar et al., 2010) (figura 1). Adicionalmente, la rápida disminución en los costos de genotipificación permite aumentar el número de animales genotipados y robustecer más los análisis a través de parentescos observados y no esperados como ocurre en la matriz de parentesco por pedigrí (Legarra, Aguilar, & Misztal, 2009). No obstante, esa tendencia podría no ser clara para países en vías de desarrollo, debido a la escasez de información genealógica, fenotípica y limitaciones económicas para el genotipado de una población grande.

### Mejor predictor lineal insesgado en una etapa (ssGBLUP)

El ssGBLUP permite, a partir de la covarianza de frecuencias génicas, modelar el genotipo de un individuo a partir de la matriz de parentesco (Legarra & Ducrocq, 2012). Esta información es incluida en el sistema clásico de ecuaciones del modelo mixto, a través de la adición de una matriz de parentesco modificada y basada en las diferencias de parentesco genómico y por pedigrí de animales con genotipos (Fragomeni et al., 2015). El sistema de ecuaciones del modelo mixto que incluye la información genómica en ssGBLUP propuesto por Misztal et al. (2009) es el siguiente:

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

Donde  $\hat{b}$  es un vector de soluciones para efectos fijos,  $\hat{a}$  es un vector de soluciones para efectos aleatorios, y  $y$  es un vector de observaciones,  $X$  es una matriz de incidencia que relaciona las observaciones con los efectos fijos,  $Z$  es una matriz de incidencia que relaciona las observaciones con el vector de efectos aleatorios genéticos aditivos directos,  $\alpha$  es el cociente entre la varianza ambiental y la varianza genética aditiva directa y  $H$  es una matriz que combina parentescos por pedigrí y genómicos en una población en la que solamente algunos individuos están genotipados. Este modelo animal tiene

la misma estructura propuesta por Henderson (1984), a excepción de la matriz de parentesco por pedigrí ( $A$ ), la cual es reemplazada por la matriz  $H$ .

Igualmente, este modelo asume que los efectos de los marcadores tienen una distribución normal con varianza común (Legarra et al., 2009). La matriz  $H$  se obtiene a partir de la siguiente operación matricial:

$$H = A + A_{\Delta}$$

Donde  $A$  es la matriz de parentesco por pedigrí y  $A_{\Delta}$  contiene las diferencias de parentesco genómicas y aditivas de los animales genotipados:

$$A_{\Delta} = \begin{bmatrix} 0 & 0 \\ 0 & G - A_{22} \end{bmatrix}$$

La estimación de la matriz genómica ( $G$ ) es generada como  $RR'/k$ , en donde  $R$  es una matriz que se obtiene a partir de  $(P-M)$ , siendo  $M$  una matriz que contiene la información de cada marcador que los individuos heredaron (VanRaden, 2008),  $P$  es una matriz que contiene las frecuencias alélicas expresadas como una diferencia desde 0,5 y multiplicado por 2, siendo cada columna de  $P$  representada por  $2(p_i - 0,5)$ ;  $k$  es un escalar calculado a partir de  $2\sum p_i(1-p_i)$ , interpretando  $p_i$  como la frecuencia del alelo de referencia para todos los casos. La sustracción  $G - A_{22}$  se realiza para evitar redundancia en la información de parentesco y la inflación en las estimaciones de los valores genómicos (Liu, Goddard, Reinhardt, & Reents, 2014; Misztal et al., 2013).

Aguilar et al. (2010) realizaron la primera evaluación genética con ssGBLUP en ganado Holstein en los Estados Unidos, obteniendo mayores precisiones. A partir de ese momento, ssGBLUP se ha convertido en un método más simple y más preciso para la estimación de valores genéticos. Sin embargo, para su implementación se debe genotipar una parte de la población, considerar exigencias computacionales y evaluar factores que podrían influir en las precisiones de las estimaciones como la cantidad de animales genotipados y su relación con la población evaluada.

## Población genotipada en ssGBLUP

En general, el tamaño de una población genotipada en programas de selección genómica debe ser mayor cuando la población candidata a selección tiene una menor relación genética con la población genotipada (Meuwissen, 2009). Asimismo, la información genómica de los ancestros es importante para aumentar las precisiones (Mulder, Calus, Druet, & Schrooten, 2012), pero en ocasiones esto es difícil debido a que la información genómica generalmente proviene de los animales más jóvenes en la población. Específicamente para ssGBLUP, la construcción de la matriz  $G$  debería ser construida con base en las frecuencias alélicas de los animales fundadores de la población evaluada (Christensen et al., 2012). Sin embargo, la construcción de esta matriz no es siempre factible en la práctica debido a que usualmente los animales fundadores no están genotipados.

Lourenco et al. (2015) sugirieron que la estrategia de genotipado en ssGBLUP debe ser dirigida hacia los animales con mayor importancia, que generalmente son los más antiguos y con mayor información genealógica y fenotípica. Consecuentemente, la composición y el tamaño de la población genotipada son factores que afectan la estructura de varianzas y covarianzas de los valores genéticos y con ello la precisión. Aunque en ocasiones las poblaciones genotipadas para ssGBLUP son pequeñas, es importante tener en cuenta que poblaciones genotipadas de mayor tamaño contribuirán más para la estimación de efectos alélicos y con una mayor precisión (VanRaden, 2016). Actualmente, un mayor número de animales genotipados con una mayor cantidad de SNP es más accesible por la disminución de los costos de la genotipificación y una mayor eficiencia en métodos estadísticos y computacionales en procesos de imputación (Uemoto, Sasaki, Sugimoto, & Watanabe, 2015).

### Aspectos computacionales

Los métodos basados en el BLUP genómico requieren la inversa de la matriz de parentesco genómica (Pocrnic, Lourenco, Mazuda, Legarra, & Misztal, 2016). El

costo de la inversión para  $G$  y  $A$  de los animales genotipados es una función cúbica del número de animales genotipados (Legarra & Ducrocq, 2012), condición que dificulta la convergencia de las evaluaciones. Experiencias con bases de datos con más de un millón de animales incluidos mostraron que la convergencia de la evaluación genética depende de la matriz  $G$  usada, la cual es generalmente singular y no puede ser invertida sin procedimientos adicionales (Aguilar et al., 2010). Por esta razón, VanRaden (2008) definió  $G=RR'/k$  con el objetivo de obtener la matriz  $G$  en una forma semidefinida positiva. A pesar de eso, la matriz aún puede ser singular cuando existen dos individuos con el mismo genotipo (gemelos idénticos), condición que causa singularidad en la matriz  $A$ . En el mismo estudio, se desarrolló un método basado en la ponderación de las matrices  $G$  y  $A$  de los animales genotipados, así:

$$wG + (1 - w)A$$

Donde  $w$  es el ponderador dado a la matriz genómica para facilitar las operaciones matriciales. Lo anterior evita problemas de singularidad y facilita la construcción de la matriz que contiene las diferencias entre parentesco por pedigrí y genómico de animales genotipados. Aunque  $w$  podría afectar las estimaciones, Aguilar et al. (2010) mostraron que valores de ponderación entre 0,95 y 0,98 no reportaron diferencias considerables y sí generó una mayor precisión de la metodología ssGBLUP frente a msGBLUP.

La genotipificación masiva de animales permitió tener un mayor número de animales genotipados a la cantidad de SNP usada en el genotipado. Esta condición impidió de nuevo la inversión convencional de la matriz genómica, debido a que en ese escenario las matrices siempre son singulares (Liu, Goddard, Hayes, Reinhardt, & Reents, 2016). A partir de esto, varias aproximaciones computacionales han sido propuestas en los últimos años, pero todas tienen problemas de convergencia o son de alto costo y complejas para su programación, en especial cuando se usan modelos multicarácter y regresiones

aleatorias (Fragomeni et al., 2015). Inicialmente, la programación de ssGBLUP incluyó las ecuaciones previamente mostradas, pero esta aplicación estaría limitada para un número aproximado de hasta cien mil animales genotipados. Una cantidad superior dificultó hallar la inversa de la matriz genómica de parentesco debido a las limitaciones en memoria y tiempo de computación (Misztal, Legarra, & Aguilar, 2014; Misztal et al., 2013).

Ese obstáculo era predecible debido a que ssGBLUP permite incluir en su metodología todos los animales en la evaluación, exigiendo algoritmos más eficientes (Liu et al., 2014). Uno de los últimos avances para ssGBLUP propuestos por Misztal et al. (2014) fue el desarrollo del algoritmo para animales probados y jóvenes (APY), que divide la población genotipada en animales con registros y animales jóvenes. Este algoritmo hace uso de la recursión genómica para aproximar la inversa de la matriz genómica.

Las recursiones genómicas se pueden estimar a partir de la adición del coeficiente de parentesco genómico de cada animal y la varianza condicional (Misztal et al., 2014). Estas estimaciones están basadas en el supuesto de que existe un limitado número de segmentos de cromosomas o efectos de SNP que explican el total de la varianza genética (Pocrnic et al., 2016). El uso de este algoritmo redujo drásticamente los costos cuando el número de animales genotipados fue mayor a cincuenta mil y ayudó a la estabilidad del proceso iterativo. Incluso, la evaluación con un millón de animales genotipados es posible (Misztal et al., 2014).

Misztal (2016) mostró que, al usar la teoría de las recursiones genómicas apoyado en el algoritmo APY, la correlación entre las estimaciones halladas con el algoritmo frente a las halladas con la inversión directa y completa de la matriz genómica fueron superiores a 0,99. Por lo tanto, el algoritmo APY facilita la evaluación genómica, estimación de parámetros genéticos y estudios de asociación genómica en grandes poblaciones genotipadas. Por ejemplo, con una población mayor a quinientos mil animales, diez mil animales fueron suficientes en la recursión

genómica para obtener estimaciones incluso más precisas que la inversión completa de la matriz genómica (Pocrnic et al., 2016).

Alternativas adicionales han incluido la implementación de un método indirecto de ssGBLUP apoyado en el algoritmo APY. Allí, los efectos de los SNP pueden ser calculados con el método convencional ssGBLUP, obteniendo un valor genómico directo. Después, este usa ponderaciones idénticas en toda la población para las diferentes fuentes de información, lo que disminuye el tiempo computacional requerido para estimar valores genéticos de animales jóvenes genotipados (Lourenco et al., 2014b). Así, una población genotipada mayor a cien mil animales podría presentar redundancia en los datos genómicos, por lo que un subconjunto de esta población y el uso de métodos indirectos podrían explicar la misma varianza genética estimada con toda la población. Aunque algunos problemas asociados a la dimensión de datos genómicos han sido solucionados, sin duda la selección genómica continuará evolucionando con los nuevos desarrollos en secuenciación y genotipificación (Garrick et al., 2014). No obstante, los algoritmos y recursos computacionales se están volviendo más limitantes que otros aspectos debido a que el número de animales y marcadores genotipados aumentan rápidamente (VanRaden, 2016).

### Ventajas de ssGBLUP

El ajuste de las (co)varianzas entre individuos realizado en ssGBLUP permite obtener una derivación automática de las ponderaciones para la combinación de diferentes fuentes de información en la estimación de los valores genéticos, siendo una forma de evaluación genética más práctica y precisa, incluso para análisis multicarácter (Aguilar et al., 2010; Lourenco et al., 2014a). Además, las estimaciones a partir de modelos unicarácter o multicarácter con ssGBLUP permiten obtener valores genéticos para animales no genotipados, lo que no es posible en análisis con msGBLUP (Tsuruta, Misztal, Aguilar, & Lawlor, 2011).

Posiblemente, la precisión en las evaluaciones genéticas es el aspecto de mayor importancia, porque permite la clasificación adecuada de los animales y un mayor progreso genético si se asume la misma intensidad de selección. En selección genómica, un mecanismo por el cual aumentan las precisiones es a través de la validación y control de calidad del pedigrí (Christensen et al., 2012). ssGBLUP permite identificar conflictos en el pedigrí y ajustar relaciones de parentesco entre animales genotipados y no genotipados que conforman los archivos genealógicos de la población evaluada, disminuyendo sesgos en el ajuste de valores genéticos para animales sin información fenotípica.

Asimismo, los valores de consanguinidad son más precisos en ssGBLUP, porque la segregación mendeliana es estimada con eventos observados (SNP) y no con probabilidades medias esperadas incluidas en la matriz de parentesco por pedigrí (De Roos, Schrooten, Veerkamp, & Arendonk, 2011). Por ejemplo, dos animales genotipados que no tienen relación en la matriz de parentesco por pedigrí podrían mostrar relación en la matriz genómica debido a un ancestro en común que no fue registrado en el pedigrí. Igualmente, es posible que aparezcan mayores coeficientes de consanguinidad en los descendientes de animales genotipados tanto como estos estén relacionados (Legarra et al., 2009). La selección genómica puede tener entonces un rol fundamental en el control de la consanguinidad de las poblaciones, lo que permite un mayor control sobre la preservación de la diversidad genética de las poblaciones (Wensch-Dorendorf, Yin, Swalve, & König, 2011), y estimar con mayor precisión el efecto negativo de la consanguinidad sobre el desempeño fenotípico.

Por otro lado, una mayor cantidad de información conduce generalmente a una mayor exactitud de las estimaciones en programas de selección. Por esta razón, las precisiones y la varianza genética aditiva también aumentan cuando se usan chips de más alta densidad (Jattawa, Elzo, Koonawootrittriron, & Suwanasopee, 2015). Sin embargo, algunas poblaciones como la reportada por Loberg et al. (2015) explican un mayor porcentaje de la varianza

genética cuando se utiliza la matriz de parentesco por pedigrí. Igualmente, Haile, Nieuwhof, Beard, Konstatinov y Hayes (2013) encontraron que la proporción genética de la varianza explicada por la información genómica y el pedigrí varió considerablemente entre características, sugiriendo que las discrepancias podrían ser dependientes de factores como la heredabilidad, e información fenotípica y genómica.

A pesar de que los chips de alta densidad no capturan toda la varianza genética, la precisión en la selección de animales jóvenes es mayor. Sin embargo, esta variación también depende de otros factores como el tamaño y la relación de parentesco de la población genotipada y no genotipada. Aunque el pedigrí en algunos casos explica una mayor proporción de la varianza, el error fue más alto en las estimaciones basadas en el pedigrí (Haile et al., 2013). Esto indica que, además de la población genotipada, los parámetros genéticos y la magnitud de sus estimaciones también podrían afectar las estimaciones genómicas, que están basadas en la (co)varianza encontrada entre los individuos de la población genotipada y su relación con animales no genotipados.

Respecto al componente económico, la reducción del intervalo generacional, el aumento en la precisión de los valores genéticos y una posible disminución en la toma de registros fenotípicos en empresas ganaderas han sido razones suficientes para justificar la implementación de programas de selección genómica. Schaeffer (2006) indicó que la selección genómica podría incluso permitir la existencia de hatos en los que la toma de registros no sea necesaria, por lo que las decisiones de selección estarían hechas con base en información genómica. Por ahora, ssGBLUP parece ser la alternativa más eficiente para ajustar simultáneamente información de animales genotipados y no genotipados, en especial para países con un número bajo de animales genotipados, en los que el acceso a la genotipificación de toda la población puede ser limitado e incluso nulo. Esto ubica a ssGBLUP como la metodología con más oportunidades e impactos económicos para reemplazar los sistemas clásicos de evaluación genética (Christensen et al., 2012).



## Limitaciones de ssGBLUP

Los métodos estadísticos basados en la distribución de los efectos de los SNP pueden ser clasificados en dos grupos: el primero asume que los SNP tienen algún efecto sobre la característica y que la varianza de cada efecto es igual; el segundo permite a los marcadores tener diferentes tipos de distribución (Zhang et al., 2010). Las predicciones genómicas lineales se rigen bajo el modelo infinitesimal, que asume que todos los SNP contribuyen uniformemente con la varianza genética, por lo que no es posible identificar loci de gran efecto asociados a características cuantitativas (Goddard, 2009). La metodología ssGBLUP no tiene en cuenta los efectos mayores de loci dado el supuesto de efecto infinitesimal con el que fue construida ssGBLUP. Aunque el supuesto de varianza común permanece en continua discusión y parecería ser una desventaja, el trabajo con otro tipo de distribución y que contemplen loci de efecto mayor no ha presentado ganancias importantes en la precisión respecto a ssGBLUP.

Legarra y Ducrocq (2012) describieron que las tres limitantes más importantes que tiene ssGBLUP son las siguientes: 1) el costo de computación, 2) la necesidad de modificar el *software* existente, que cubre una variedad de modelos usados de evaluación, y 3) su forma lineal, que excluye los estimadores no lineales incluyendo información genómica. Aunque los desarrollos científicos y computacionales indicados en esta revisión han cubierto las dos primeras problemáticas, la discusión sigue presente en los supuestos que asume el modelo. Estudios con simulación han mostrado que las aproximaciones bayesianas han resultado en estimaciones de mayor precisión; sin embargo, evaluaciones con datos reales en producción de leche han terminado en algunas ocasiones con estimaciones de menor precisión (Su et al., 2012). Además, Aguilar et al. (2010) mostraron que al usar un número grande de SNP con igual varianza fue apropiado para la mayoría de las características productivas.

A pesar de que el escenario más realista resulta cuando el modelo asume que existen genes de efecto

mayor, estudios en los que se compararon las estimaciones genómicas a través de metodologías bayesianas y ssGBLUP mostraron que las precisiones y las pérdidas de varianza son muy similares (Wang, Misztal, Aguilar, Legarra, & Muir, 2012). Por lo tanto, ssGLUP tendría una mayor ventaja frente a otras metodologías que no permiten la inclusión de animales no genotipados y requieren datos fenotípicos para la derregresión de los valores genéticos. Esto limita la aplicación práctica de la selección genómica en poblaciones en las que no es posible genotipar toda la población, condición que a menudo se presenta. Sin embargo, cuando el número de registros es bajo y la densidad de los chips usados es alta, otras metodologías parecen ser superiores a ssGBLUP (Lourenco et al., 2014b).

## Uso de ssGBLUP en poblaciones pequeñas

El interés de la selección genómica es identificar un conjunto de SNP para incrementar la precisión de los valores genéticos. Para esto, es necesario estimar los efectos de los SNP a partir de la población genotipada (Misztal, 2016; Misztal & Legarra, 2016). Una de las alternativas más útiles que permite económicamente incrementar el número de animales genotipados han sido los genotipos imputados desde chips de varias densidades y que son combinados en una sola evaluación genómica para reducir los costos y aumentar la precisión. Por ejemplo, el número de animales genotipados en los Estados Unidos de la raza Holstein incrementó considerablemente a más de 950.000 animales en 2016 (Misztal, 2016).

La precisión de la imputación y las estimaciones genómicas también se ven afectadas cuando el número de animales genotipados, el desequilibrio de ligamiento ( $<0,2$ ) y las relaciones genéticas entre las poblaciones genotipadas son bajas (Jattawa et al., 2015; Vanraden et al., 2013), debido a que no se pueden estimar con precisión los segmentos de haplotipos compartidos entre los animales relacionados y consecuentemente predecir los genotipos de los otros SNP de animales genotipados con chips de más baja densidad.

Un aumento del número de marcadores para los animales genotipados también podría ser una estrategia; no obstante, esto tuvo efectos contradictorios en las precisiones. Al aumentar el número de SNP, la tasa de recombinación disminuye, razón por la cual la precisión aumenta. Sin embargo, un aumento en el número de marcadores sin un aumento en el número de fenotipos es contraproducente, debido a que la colinealidad podría confundir los efectos y disminuir la precisión (Muir, 2007). Esta alternativa tendría incluso mayores contradicciones en países con poblaciones limitadas, ya que generalmente la toma de registros fenotípicos también es escasa.

La construcción de estrategias para mejorar la predicción en poblaciones pequeñas debe surgir a partir de los conceptos teóricos de ssGBLUP para incrementar la confiabilidad. Estos requerimientos son los siguientes: 1) más genotipos debe resultar en igual o mayor confiabilidad; 2) un animal joven que ha sido genotipado no debe crear información adicional para otros animales; 3) la información extra de que un animal joven con ancestros no genotipados pueda contribuir a la población genotipada debe ser pequeña o nula, y 4) valores adicionales de confiabilidad no pueden ser obtenidos a partir de animales de líneas o razas diferentes (Misztal et al., 2013).

Los países con poblaciones pequeñas y con ausencia de evaluaciones genéticas generalmente soportan programas de selección en la continua importación de material genético de poblaciones mejoradas (Andonov et al., 2016). En selección genómica, esto es más importante cuando la correlación del desempeño fenotípico entre diferentes países es alta (Vanraden, 2016), indicando que las estimaciones genómicas para animales provenientes de poblaciones pequeñas pueden ser derivados de los efectos de los SNP estimados en otras poblaciones genotipadas (Andonov et al., 2016), y el ajuste de la matriz de parentesco podría ser mayor dadas las contribuciones genéticas de toros usados en programas de inseminación artificial en países con poblaciones pequeñas.

El intercambio de información genómica para incrementar el tamaño de la población genotipada es una aproximación eficiente para incrementar la

precisión de las estimaciones cuando el número de animales genotipados es bajo (Wiggans et al., 2015). No obstante, cuando la población contiene animales locales e importados, registros fenotípicos son aún necesarios para aumentar la precisión de la progenie (Lourenco et al., 2014a). Un camino para fortalecer esta alternativa es la inclusión de hembras con alta contribución genética y sus desempeños fenotípicos (Mc Hugh, Meuwissen, Cromie, & Sonesson, 2011).

La inclusión de hembras genotipadas es deseable porque son parte importante de los programas de mejoramiento al estar sometidas igualmente a un proceso de selección y aportar valores fenotípicos a través de su propio desempeño y su progenie. Además, permiten incrementar el tamaño de la población genotipada, convirtiéndolas en una alternativa para reducir los sesgos y aumentar la precisión (Tsuruta, Misztal, & Lawlor, 2013). Asimismo, la inclusión de hembras podría ajustar con mayor precisión los grados de parentesco a partir de los efectos de los marcadores (Loberg et al., 2015). A pesar de que Tsuruta et al. (2013) evaluaron la inclusión de hembras en una población genotipada que no tenía limitaciones en tamaño, un criterio de selección fue la inclusión de vacas con registros y valores genéticos similares a los toros incluidos. Por lo tanto, la inclusión de vacas en poblaciones pequeñas y con escasez de evaluaciones genéticas podría mejorar considerablemente la implementación de un programa de selección genómica a través de la metodología ssGBLUP.

## Conclusiones

La metodología ssGBLUP es una aproximación práctica y más accesible para incluir información genómica a través de una mayor exactitud de la (co) varianza genética entre individuos en programas de selección. Aunque las limitaciones más importantes de ssGBLUP parecen estar dirigidas hacia aspectos computacionales, los continuos desarrollos en programación han permitido el desarrollo e implementación de algoritmos para las estimaciones genómicas. Finalmente, es necesario incluir información genómica

de hembras locales y toros extranjeros que permitan aumentar la conectividad genética de la población genotipada y no genotipada para obtener mejores estimaciones de los valores genéticos.

## Agradecimientos

A la Asociación Colombiana de Criadores de Ganado Simmental (Asosimmental) y al departamento de genética de AGROSAVIA en Tibaitatá en el marco del convenio TV16-11 de cooperación para llevar a cabo el proyecto “Sistema de información basado en el desarrollo de pruebas de desempeño en crecimiento para la identificación de reproductores de alto mérito genético en la raza Simmental y Simbrah”. A la convocatoria de sostenibilidad de la Universidad

de Antioquia para el grupo de investigación GaMMA. El primer autor agradece a la Universidad de Ciencias Aplicadas y Ambientales (UDCA) y a Colciencias (convocatoria 727 de 2015) por el apoyo financiero para su formación en el doctorado en Ciencias Animales de la Universidad de Antioquia.

## Descargos de responsabilidad

El manuscrito fue ejecutado, preparado y revisado con la participación de todos los autores, quienes declaramos que no existe conflicto de intereses que ponga en riesgo la validez de los resultados presentados.

## Referencias

- Aguilar, I., Misztal, I., Johnson D., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*(2), 743-752. doi:10.3168/jds.2009-2730.
- Andonov, S., Lourenco, D. A. L., Fragomeni, B. O., Masuda, Y., Pocrnic, I., Tsuruta, S., & Misztal, I. (2016). Accuracy of breeding values in small genotyped populations using different sources of external information-a simulation study. *Journal of Dairy Science*, *100*(1), 395-401. doi:10.3168/jds.2016-11335.
- Chen, J., Wang, Y., Zhang, Y., Sun, D., Zhang, S., & Zhang, Y. (2011). Evaluation of breeding programs combining genomic information in Chinese Holstein. *Agricultural Sciences in China*, *10*(12), 1949-1957. doi:10.1016/S1671-2927(11)60196-X.
- Christensen, O., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, *6*(10), 1565-1571. doi:10.1017/S1751731112000742.
- De Roos, A. P. W., Schrooten, C., Veerkamp, R. F., & Van Arendonk, J. A. M. (2011). Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *Journal of Dairy Science*, *94*(3), 1559-1567. doi:10.3168/jds.2010-3354.
- Elzo, M. A., Thomas, M. G., Johnson, D. D., Martinez, C. A., Lamb, G. C., Rae, D. O., & Driver, J. D. (2015). Genetic parameters and predictions for direct and maternal growth traits in a multibreed Angus-Brahman cattle population using genomic-polygenic and polygenic models. *Livestock Science*, *178*, 43-51. doi:10.1016/j.livsci.2015.06.015.
- Fragomeni, B. O., Lourenco, D. A. L., Tsuruta, S., Masuda, Y., Aguilar, I., Legarra, A., & Misztal, I. (2015). Hot topic: Use of genomic recursions in single-step genomic Best Linear Unbiased Predictor (BLUP) with a large number of genotypes. *Journal of Dairy Science*, *98*(6), 4090-4094. doi:10.3168/jds.2014-9125.
- Garrick, D., Dekkers, J., & Fernando, R. (2014). The evolution of methodologies for genomic prediction. *Livestock Science*, *166*(1), 10-18. doi:10.1016/j.livsci.2014.05.031.
- Garrick, D., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, *31*, 41-55. doi:10.1186/1297-9686-41-55.
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long-term response. *Genetica*, *136*(2), 245-257. doi:10.1007/s10709-008-9308-0.
- Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews*, *10*(6), 381-391. doi:10.1038/nrg2575.
- Haile, M., Nieuwhof, G. J., Beard, K. T., Konstatinov, K. V., & Hayes, B. J. (2013). Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. *Journal of Animal Breeding and Genetics*, *130*(1), 20-31. doi:10.1111/j.1439-0388.2013.01001.x.
- Henderson, C. R. (1984). *Applications of linear models in animal breeding*. Guelph, Canadá: University of Guelph.
- Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *Genes Genomes Genetics*, *4*(6), 1027-1046. doi:10.1534/g3.114.010298.

- Jattawa, D., Elzo, M. A., Koonawootrittriron, S., & Suwanasopee, T. (2015). Comparison of genetic evaluations for milk yield and fat yield using a polygenic model and three genomic-polygenic models with different sets of SNP genotypes in Thai multibreed dairy cattle. *Livestock Science*, *181*, 58-64. doi:10.1016/j.livsci.2015.10.008.
- König, S., & Swalve, H. H. (2009). Application of selection index calculations to determine selection strategies in genomic breeding programs. *Journal of Dairy Science*, *92*(10), 5292-5303. doi:10.3168/jds.2009-2232.
- Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, *92*(9), 4656-4663. doi:10.3168/jds.2009-2061.
- Legarra, A., & Ducrocq, V. (2012). Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. *Journal of Dairy Science*, *95*(8), 4629-4645. doi:10.3168/jds.2011-4982.
- Liu, Z., Goddard, M. E., Hayes, B. J., Reinhardt, F., & Reents, R. (2016). Technical note: Equivalent genomic models with a residual polygenic effect. *Journal of Dairy Science*, *99*(3), 2016-2025. doi:10.3168/jds.2015-10394.
- Liu, Z., Goddard, M. E., Reinhardt, F., & Reents, R. (2014). A single-step genomic model with direct estimation of marker effects. *Journal of Dairy Science*, *97*(9), 5833-5850. doi:10.3168/jds.2014-7924.
- Loberg, A., Durr, J. W., Fikse, W. F., Jorjani, H., & Crooks, L. (2015). Estimates of genetic variance and variance of predicted genetic merits using pedigree or genomic relationship matrices in six Brown Swiss cattle populations for different traits. *Journal of Animal Breeding and Genetics*, *132*(5), 376-385. doi:10.1111/jbg.12142.
- Lourenco, D. A. L., Misztal, I., Tsuruta, S., Aguilar, I., Ezra, E., Ron, M., & Weller, J. I. (2014a). Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. *Journal of Dairy Science*, *97*(3), 1742-1752. doi:10.3168/jds.2013-6916.
- Lourenco, D. A. L., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T. J., Forni, S., & Weller, J. I. (2014b). Are evaluations on young genotyped animals benefiting from the past generations? *Journal of Dairy Science*, *97*(6), 3930-3942. doi:10.3168/jds.2013-7769.
- Lourenco, D. A. L., Tsuruta, S., Fragomeni, B. O., Masuda, Y., Aguilar, I., Legarra, A., & Misztal, I. (2015). Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *Journal of Dairy Science*, *93*(6), 2653-2662. doi:10.2527/jas.2014-8836.
- Mc Hugh, N., Meuwissen, T. H. E., Cromie, C. R., & Sonesson, A. K. (2011). Use of female information in dairy cattle genomic breeding programs. *Journal of Dairy Science*, *94*(8), 4109-4118. doi:10.3168/jds.2010-4016.
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. *Genetics Selection Evolution*, *41*, 35. doi:10.1186/1297-9686-41-35.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829.
- Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics*, *202*(2), 401-409. doi:10.1534/genetics.115.182089.
- Misztal, I., & Legarra, A. (2016). Invited review: Efficient computation strategies in genomic selection. *Animal*, *11*(5), 731-736. doi:10.1017/S1751731116002366.
- Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, *92*(9), 4648-4655. doi:10.3168/jds.2009-2064.
- Misztal, I., Legarra, A., & Aguilar, I. (2014). Using recursion to compute the inverse of the genomic relationship matrix. *Journal of Dairy Science*, *97*(6), 3943-3952. doi:10.3168/jds.2013-7752.
- Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., Vanraden, P. M., & Lawlor, T. J. (2013). Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science*, *96*(1), 647-654. doi:10.3168/jds.2012-5656.
- Moser, G., Khatkar, M. S., Hayes, B. J., & Raadsma, H. W. (2010). Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genetics Selection Evolution*, *42*(1), 37. doi: 10.1186/1297-9686-42-37.
- Muir, W. M. (2007). Comparison of genomic and traditional blup-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics*, *124*(6), 342-355. doi:10.1111/j.1439-0388.2007.00700.x
- Mulder, H. A., Calus, M. P. L., Druet, T., & Schrooten, C. (2012). Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *Journal of Dairy Science*, *95*(2), 876-889. doi:10.3168/jds.2011-4490.
- Patry, C., & Ducrocq, V. (2011). Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *Journal of Dairy Science*, *94*(2), 1011-1020. doi:10.3168/jds.2010-3804.
- Pintus, M., Gaspa, G., Nicolazzi, E., Vicario, D., Rossoni, A., Ajmone-Marsan, P., & Macciotta, N. P. (2012). Prediction of genomic breeding values for dairy traits in Italian Brown and Simmental bulls using a principal component approach. *Journal of Dairy Science*, *95*(6), 3390-3400. doi:10.3168/jds.2011-4274.
- Pocrnic, I., Lourenco, D. A. L., Masuda, Y., Legarra, A., & Misztal, I. (2016). The dimensionality of genomic information and its effect on genomic prediction. *Genetics*, *203*(1), 573-581. doi:10.1534/genetics.116.187013.
- Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, *123*(4), 218-223. doi:10.1111/j.1439-0388.2006.00595.x.

- Su, G., Madsen, P., Nielsen, U. S., Mäntysaari, E. A., Aamand, G. P., Christensen, O. F., & Lund, M. S. (2012). genomic prediction for Nordic Red cattle using one-step and selection index blending. *Journal of Dairy Science*, *95*(2), 909-917. doi:10.3168/jds.2011-4804.
- Tsuruta, S., Misztal, I., Aguilar, I., & Lawlor, T. J. (2011). Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *Journal of Dairy Science*, *94*(8), 4198-4204. doi:10.3168/jds.2011-4256.
- Tsuruta, S., Misztal, I., & Lawlor, T. J. (2013). Short communication: genomic evaluations of final score for US Holsteins benefit from the inclusion of genotypes on cows. *Journal of Dairy Science*, *96*(5), 3332-3335. doi:10.3168/jds.2012-6272.
- Uemoto, Y., Sasaki, S., Sugimoto, Y., & Watanabe, T. (2015). Accuracy of high-density genotype imputation in Japanese Black cattle. *Animal Genetics*, *46*(4), 388-394. doi:10.1111/age.12314.
- Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414-4423. doi:10.3168/jds.2007-0980.
- Vanraden, P. M. (2016). Practical implications for genetic modeling in the genomics era. *Journal of Dairy Science*, *99*(3), 2405-2412. doi:10.3168/jds.2015-10038.
- Vanraden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., & Doak, G. A. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science*, *96*(1), 668-678. doi:10.3168/jds.2012-5702.
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, *94*(2), 73-83. doi:10.1017/S0016672312000274.
- Wensch-Dorendorf, M., Yin, T., Swalve, H. H., & König, S. (2011). Optimal strategies for the use of genomic selection in dairy cattle breeding programs. *Journal of Dairy Science*, *94*(8), 4140-4151. doi:10.3168/jds.2010-4101.
- Wiggans, G. R., Su, G., Cooper, T. A., & Nielsen, U. S., Aamand, G. P., Guldbbrandtsen, B., & Vanraden, P. M. (2015). Short communication: Improving accuracy of Jersey genomic evaluations in the United States and Denmark by sharing reference population bulls. *Journal of Dairy Science*, *98*(5), 3508-3513. doi:10.3168/jds.2014-8874.
- Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D. J., & Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PlosONE*, *5*(9), 1-8. doi:10.1371/journal.pone.0012648.