



UNIVERSIDAD AUTÓNOMA CHAPINGO

DEPARTAMENTO DE ENSEÑANZA, INVESTIGACIÓN Y SERVICIO EN
ZOOTECNIA

POSGRADO EN PRODUCCIÓN ANIMAL

MAESTRÍA EN CIENCIAS EN INNOVACIÓN GANADERA

EXACTITUD DE VALORES GENÓMICOS OBTENIDOS MEDIANTE DOS
MODELOS DE ANÁLISIS UTILIZANDO INFORMACIÓN SIMULADA

TESIS

Que como requisito parcial para obtener el grado de:

Maestro en Ciencias en Innovación Ganadera

Presenta:

Fernanda Ramírez Flores

Bajo la supervisión de:

Agustín Ruíz Flores, Ph. D. y Joel Domínguez Viveros, Dr.



DIRECCIÓN GENERAL ACADÉMICA
DEPARTAMENTO DE SERVICIOS ESCOLARES
OFICINA DE EXÁMENES PROFESIONALES

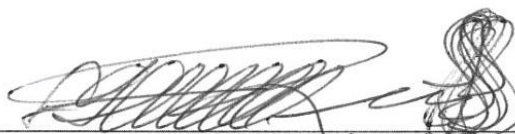


CHAPINGO, MÉXICO, JUNIO 2017

EXACTITUD DE VALORES GENÓMICOS OBTENIDOS MEDIANTE DOS MODELOS DE ANÁLISIS UTILIZANDO INFORMACIÓN SIMULADA

Tesis realizada por la Ing. Fernanda Ramírez Flores bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de: **MAESTRO EN CIENCIAS EN INNOVACIÓN GANADERA**

DIRECTOR



Ph. D. AGUSTÍN RUIZ FLORES

CO-DIRECTOR



Dr. JOEL DOMÍNGUEZ VIVEROS

ASESOR



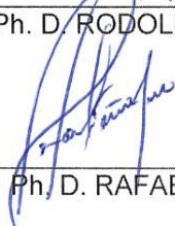
Ph. D. BALBOMERO ALARCÓN ZÚÑIGA

ASESOR



Ph. D. RODOLFO RAMÍREZ VALVERDE

ASESOR



Ph. D. RAFAEL NÚÑEZ DOMÍNGUEZ

Chapingo, Méx., abril de 2017

AGRADECIMIENTOS

A mi “Alma Mater” la Universidad Autónoma Chapingo

Al Posgrado en Producción Animal

Al Ph. D. Agustín Ruíz Flores, por ser mi asesor, confiar en mí y darme su total apoyo durante mi estadía en la maestría y para la realización de la presente tesis.

A mis asesores el Dr. Joel Domínguez Viveros, el Ph. D. Baldomero Alarcón Zúñiga, el Ph. D. Rodolfo Ramírez Valverde y el Ph. D. Rafael Núñez Domínguez por su valiosa colaboración al revisar este trabajo de tesis.

Al Consejo Nacional de Ciencia y Tecnología.

DEDICATORIAS

A mis padres, Tomás Ramírez Lucas y Marcelina Flores Cortés pilares fundamentales en mi vida, con mucho amor, les dedico todo mi esfuerzo, en reconocimiento a todo el sacrificio y la confianza puestos para llegar a este punto de mi vida, se merecen esto y mucho más.

A mis hermanos María Luisa y Ricardo que a pesar de la distancia siempre estuvieron conmigo, gracias por su sacrificio y apoyo, los quiero mucho.

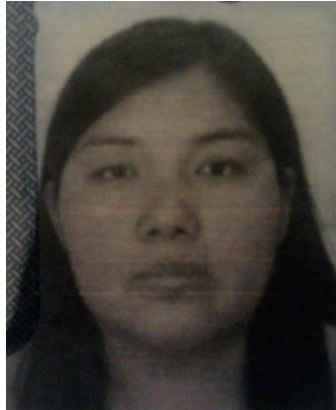
A mi hermano, hijo de otra madre, Emanuel Cobian Tovar (†). Gracias por todo lo que me enseñaste.

A mis amigas Sandra Elia Ramírez Rangel, Nidia Guadalupe Romero Espejel. Por estar ahí siempre que lo necesite, gracias por su amistad y cariño.

A Gabriela Pérez Hernández, gracias por estar siempre conmigo dándome todo tu apoyo.

FER RMZ

DATOS BIOGRÁFICOS



Datos personales

Nombre: Fernanda Ramírez Flores

Fecha de nacimiento: 4 de diciembre de 1992

Lugar de nacimiento: Tepatlaxco de Hidalgo, Puebla

CURP: RAFF921204MPLMLR07

Profesión: Ingeniero Agrónomo Especialista en Zootecnia

Cédula profesional: 10135383

Desarrollo académico

Preparatoria: Preparatoria Agrícola, Universidad Autónoma Chapingo

Licenciatura: Departamento de Zootecnia, Universidad Autónoma Chapingo

CONTENIDO

AGRADECIMIENTOS.....	iii
DEDICATORIAS.....	iv
DATOS BIOGRÁFICOS.....	v
1 INTRODUCCIÓN GENERAL	1
2 REVISIÓN DE LITERATURA	3
2.1 Selección genómica.....	3
2.1.1 Antecedentes de selección genómica	4
2.1.2 Métodos para la selección genómica	5
2.1.3 Selección genómica con paneles de diferente densidad de marcadores SNP	7
2.1.4 Selección genómica a través de especies.....	8
2.1.5 Programas de mejoramiento genético con selección genómica.....	10
2.2 Desequilibrio de ligamiento.....	11
2.2.1 Causas de desequilibrio de ligamiento	13
2.2.2 Implicaciones de desequilibrio de ligamiento.....	13
2.3 Exactitud de valores genómicos	14
2.3.1 Factores que afectan la exactitud de los valores genómicos	16
2.4 Simulación de datos en selección genómica	19
2.4.1 Implicaciones.....	20
2.4.2 Hallazgos.....	20

2.5	Literatura citada	21
3	ACCURACY OF GENOMIC VALUES PREDICTED USING DEREGRESSED PREDICTED BREEDING VALUES AS RESPONSE VARIABLES	31
3.1	Abstract.....	31
3.2	Introduction	32
3.3	Material and Methods	33
3.3.1	Data simulation.....	33
3.3.2	Response variables	34
3.3.3	Genomic analysis	34
3.4	Results and Discussion.....	35
3.4.1	Accuracy of genomic values	35
3.4.2	Genetic correlations between true breeding values and genomic values	39
3.5	Conclusions	40
3.6	Acknowledgements.....	41
3.7	Cited Literature	41

1 INTRODUCCIÓN GENERAL

En mejoramiento genético animal los individuos se seleccionan con base en predicciones de su mérito genético, las que deben tener exactitudes altas y preferentemente obtenidas a temprana edad para que el progreso genético sea rápido. Las predicciones de valores genéticos obtenidas con herramientas cuantitativas, basadas en el pedigrí se han utilizado con éxito por varias décadas para seleccionar animales. Actualmente existen herramientas tecnológicas adicionales que pueden acelerar el progreso genético. Los marcadores moleculares permiten describir el genoma de los individuos con un gran número de *loci*. Lo anterior abre posibilidades para obtener una predicción más precisa de los valores genéticos a temprana edad de los individuos (De los Campos, Hickey, Pong-Wong, Daetwyler & Calus, 2013).

Soller (1978) fue de los primeros autores en plantear el uso de información molecular para seleccionar bovinos productores de leche, sugiriendo preseleccionar animales antes de la prueba de progenie con base en información molecular. Posteriormente, Fernando & Grossman (1989) aplicaron predicción lineal insesgada mediante un modelo lineal mixto con efectos aditivos para alelos en *loci* de características cuantitativas de interés económico (MQTL) y con efectos aditivos para el resto de *loci* de características cuantitativas (QTL). Estos autores desarrollaron un algoritmo recursivo para obtener una matriz de covarianzas de efectos de los alelos MQTL. Este método incluye un algoritmo para obtener la inversa de esta matriz y permite la evaluación simultánea de efectos fijos, de MQTL y del resto de los alelos de los QTLs, utilizando las relaciones conocidas y la información de fenotipos y marcadores.

La selección genómica (SG) puede realizarse en dos etapas (Meuwissen, Hayes & Goddard, 2016), en la primera los genotipos y fenotipos de una población de entrenamiento se usan para obtener ecuaciones que posteriormente se utilizarán

en una población en evaluación para obtener los valores genómicos (Garrick, 2011). Existe otra metodología que combina pedigrí y relaciones genómicas (Legarra, Aguilar & Misztal, 2009). Este es un proceso de un solo paso que permite una evaluación sencilla que incluye información fenotípica, de pedigrí y genómica disponible y no tiene limitaciones en los modelos para el análisis (Aguilar, Misztal, Legarra & Tsuruta, 2010; Chen, Misztal, Aguilar, Legarra & Muir, 2011).

2 REVISIÓN DE LITERATURA

2.1 Selección genómica

Para Goddard & Hayes (2007), SG es una forma de selección asistida por marcadores en la que éstos cubren el genoma completo, asumiendo que todos los QTL estén en desequilibrio de ligamiento con al menos un marcador. La SG ha sido posible gracias al gran número de polimorfismos de nucleótido simple (SNP) abarcando los genomas completos de varias especies domésticas, y los métodos nuevos para determinar los genotipos en forma eficiente de un gran número de marcadores. Sin embargo, el impacto de SG en los programas de mejoramiento genético ha sido menor de lo previsto inicialmente (Dekkers, 2004). Varios factores han contribuido a esto: primero, con algunas excepciones, la proporción de la varianza explicada por los QTLs ha sido pequeña; y segundo, los recursos financieros necesarios para desarrollar las poblaciones necesarias para mapear QTL han sido considerables, limitando la amplia adopción de esta tecnología (De los Campos et al., 2013). Esto a su vez ha impulsado el desarrollo de metodologías de evaluación genómica. Lo anterior es porque generalmente se requiere estimar un gran número de parámetros con un número limitado de genotipos (De los Campos et al., 2013).

En SG, el uso de marcadores moleculares que cubran el genoma completo, teóricamente permitiría considerar la mayor parte de la variabilidad genética, para predecir el valor genético y así seleccionar a los mejores candidatos para una característica cuantitativa de interés (Meuwissen, Hayes & Goddard, 2001). Los predictores obtenidos de esta manera, se denominan valores genómicos. La SG puede realizarse mediante la estimación de los efectos de los SNP individuales o mediante el uso del mejor predictor lineal insesgado (BLUP) con una matriz de relaciones genómicas, entre otros métodos (De los Campos et al., 2013; VanRaden et al., 2009). En SG se asume que todos los marcadores estarían unidos a un gen con efectos para una característica, concentrándose en la

estimación de estos efectos, más que en probar su significancia (Meuwissen et al., 2016).

Entre las ventajas de la SG se pueden citar: reducción del intervalo entre generaciones; incremento de la tasa de progreso genético; mayores exactitudes de los valores genéticos predichos; mejora en la exactitud de valores genómicos de animales jóvenes, especialmente para aquellas características de baja heredabilidad, limitadas por el sexo, medidas tarde en la vida de los animales, costosas de medir, o que requieren retar a los animales, como en la selección para resistencia a enfermedades (Garrick, 2011; Garrick & Saatchi, 2011).

Cuando SG se realiza en dos etapas, a los animales de una población de entrenamiento (PE_n) se les determina su genotipo y se les mide su fenotipo, y se evalúan para una determinada característica con el fin de estimar los efectos de los SNP. Lo siguiente es determinar los genotipos de los candidatos a selección, población de evaluación (PE_v); al combinar la información de sus genotipos con los efectos estimados en la primera fase, es posible obtener los valores genéticos genómicos (VG_n). Con SG no se requiere el pedigrí, esencial con el uso de métodos cuantitativos tradicionales basados en modelos mixtos, y los animales seleccionados a través de SG no necesitan ser evaluados fenotípicamente para la característica de interés (Meuwissen et al., 2016).

2.1.1 Antecedentes de selección genómica

En décadas recientes se han logrado avances sustanciales mediante la aplicación de técnicas de genética molecular en la identificación de *loci* y regiones cromosómicas que contienen *loci* que influyen en características cuantitativas de importancia en producción animal (Andersson, 2001). Lo anterior ha permitido mejorar las oportunidades de programas de mejoramiento genético en el ganado, mediante la selección directa de genes o regiones genómicas que afectan

características económicas, a través de la selección asistida por marcadores e introgresión de genes (Dekkers & Hospital, 2002). Con este fin, se han realizado estudios teóricos para determinar la mejor manera de utilizar la información genética molecular en programas de selección.

Las respuestas a la selección predichas en varios estudios, inicialmente dieron lugar a un gran optimismo por el uso de información de genética molecular en el mejoramiento genético (Dekkers, 2004). Esto ha sido especialmente valioso y utilizado en ganado bovino lechero, especie en la que primero se dieron las condiciones para llevar a cabo SG (Ibanez-Escriche & Simianer, 2015). De acuerdo con Meuwissen et al. (2016), tres condiciones han permitido el amplio uso de información del ADN: 1) la metodología para realizar SG (Meuwissen et al., 2001), 2) la identificación de miles de marcadores SNP, y 3) las tecnologías de bajo costo para la determinación eficiente de genotipos en términos de SNP.

2.1.2 Métodos para la selección genómica

GBLUP (Genomic Best Linear Unibased Predictor)

Con las metodologías de modelos mixtos, utilizando el BLUP, los VG se predicen utilizando los fenotipos y las relaciones familiares basadas en el pedigrí de los animales. Con el método GBLUP, los VGn se predicen utilizando los fenotipos y las relaciones genómicas, basadas en información densa de marcadores cubriendo el genoma completo. El método GBLUP es similar al BLUP, excepto que las relaciones de pedigrí se sustituyen por las genómicas. Una ventaja práctica del GBLUP es que se pueden utilizar tanto la metodología del BLUP como el software desarrollados para su obtención, sólo tiene que sustituirse el pedigrí por las relaciones genómicas (Meuwissen et al., 2016).

El GBLUP es más exacto que el BLUP porque las relaciones genómicas son más precisas que las basadas en el pedigrí (Meuwissen et al., 2016). Sin embargo,

una desventaja del uso de información del ADN, es que los predictores genómicos deberán estar basados en un número suficientemente grande de SNPs. En bovinos se ha estimado que se requieren alrededor de 50,000 SNPs distribuidos a lo largo del genoma completo (Goddard, Hayes & Meuwissen, 2011).

Imputación de genotipos faltantes

Después de determinar los genotipos mediante un chip-ADN, algunos genotipos para algunos *loci* no se conocerán. Esto puede solucionarse mediante la imputación de genotipos. Con base en los genotipos conocidos de los animales, el haplotipo que lleva el animal es reconocido, ya que el mismo fue observado en otros animales. Así, el genotipo faltante se puede encontrar en el genotipo de los animales que llevan el mismo haplotipo. Algunos programas desarrollados para la imputación de genotipos faltantes son: Beagle (Browning & Dorado, 2007), Fimpute (Sargolzaei, Chesnais & Schenkel, 2014) y Alphaimpute (Hickey; Crossa, Babu & De los Campos, 2012).

Los métodos de imputación pueden utilizarse como una herramienta estratégica para reducir costos, mediante la combinación de paneles de SNP de alta y baja densidad. A los ancestros de alto valor económico se les determina su genotipo con un chip de SNP de alta densidad y costoso, para identificar los haplotipos de la población. Posteriormente, a un gran número de descendientes se les determina su genotipo con un chip de SNP de baja densidad, de menor costo. El chip de baja densidad tiene suficientes SNPs para reconocer los haplotipos que tiene el animal y los que falten pueden ser imputados, puesto que se conocen los de sus progenitores (Meuwissen et al., 2016; Nieuwhof, Beard, Konstantinov, Bowman & Hayes, 2010). Después de la imputación se realiza la evaluación genómica.

Animales no genotipificados

Generalmente en SG, a muchos animales no se les determina su genotipo, pero se necesita incluir su información fenotípica en la predicción de los VG, tal como se hace con las metodologías cuantitativas tradicionales. Una forma de hacerlo con SG considera varios pasos: 1) para los animales a los que se les determinó su genotipo, se estiman pseudo-fenotipos que incluyen información de sus parientes no genotipificados, 2) la predicción genómica se realiza usando los pseudo-registros y sus genotipos, y 3) los VGn y VG tradicionales se combinan en un VG predicho total (Meuwissen et al., 2016).

Para la predicción genómica se ha derivado una extensión de un modelo mixto lineal a animales no genotipados. La extensión del método permite integrar de manera óptima la información genómica, de pedigrí y fenotipo en un procedimiento de una sola etapa para la estimación del valor genómico. Debido a la simplicidad del método, tal procedimiento de un solo paso tiene el potencial de convertirse en la herramienta estándar para la predicción genómica de las evaluaciones prácticas de ganado en el futuro (Christensen & Lund, 2010).

Por ejemplo, se puede utilizar la producción promedio de las hijas de un semental lechero. Aquí, al toro se le determina su genotipo pero no su fenotipo, mientras que a sus hijas se les mide el fenotipo pero podría no determinárseles su genotipo. Puesto que los datos se manejan en varios pasos, este método es claramente no el óptimo. Sin embargo, en la práctica, se han obtenido exactitudes relativamente altas con este método de SG (VanRaden et al., 2009).

2.1.3 Selección genómica con paneles de diferente densidad de marcadores SNP

En porcicultura, la estrategia de determinar los genotipos utilizando paneles con gran densidad de marcadores SNP a algunos individuos y de baja densidad a

otros, puede ser un enfoque rentable para la implementación de SG en esta especie, la desventaja radica en una pequeña pérdida de información (Habier, Fernando & Dekkers, 2009).

Los VGn se pueden predecir en cerdos utilizando genotipos imputados de un panel de marcadores de baja densidad, pero la exactitud de los VG dependerá del nivel de genotipificado en parientes cercanos, debido a su impacto en la precisión de la imputación y el tamaño del conjunto de datos. La aplicación práctica de la selección genómica en poblaciones comerciales de cerdos requiere de estrategias para los animales genotipificados con paneles de densidad alta y baja de marcadores, debido al gran número de candidatos para la selección y las limitaciones en la cantidad de genotipificaciones posibles (Cleveland & Hickey, 2013).

2.1.4 Selección genómica a través de especies

La implementación de SG en el ganado bovino lechero se ha traducido en un aumento de la ganancia genética, demostrado por análisis de tendencias genéticas en varios países. Por ejemplo, en Canadá, la tasa de ganancia genética aproximadamente se ha duplicado con la implementación de SG (Meuwissen et al., 2016).

Con SG se ha logrado una mayor confiabilidad en la predicción de VGn en el ganado Holstein mediante el panel Illumina Bovino SNP50 BeadChip en los Estados Unidos, Australia, los Países Bajos y Nueva Zelanda (Hayes, Bowman, Chamberlain & Goddard, 2009; VanRaden et al., 2009), y la expansión de SG a múltiples razas lecheras está en curso utilizando el chip Illumina BovineHD.

Los bovinos lecheros fue la primera especie en la que se cumplieron las condiciones necesarias para la implementación de la SG mencionadas previamente (Ibanez-Escriche & Simianer, 2015; Meuwissen et al., 2016). La cooperación internacional en el establecimiento de poblaciones de referencia ha

sido clave para lograr los avances de SG en esta especie. Sin embargo, Schenkel (2012) sugiere que SG ha incrementado la tasa anual de incremento en consanguinidad. Por tanto, la tasa de ganancia genética deberá maximizarse a la vez que se restrinja la del incremento en consanguinidad (Meuwissen et al., 2016).

En contraste con la industria lechera, la industria de ganado bovino productor de carne se caracteriza por poblaciones de diversa composición y mezcla de razas. Las predicciones genómicas para este ganado podrían desarrollarse para cada raza de forma independiente, pero se necesitarían grandes poblaciones de entrenamiento a desarrollar para cada raza por separado, lo que puede ser difícil para razas con poblaciones pequeñas y las diferencias entre estas para la predicción genómica en animales cruzados (Weber et al., 2012).

Existen casos de SG hecha en bovinos productores de carne y las exactitudes que se obtienen son menores que las obtenidas en bovinos lecheros (Meuwissen et al., 2016), variando de 0.3 a 0.7 (Van Eenennaam, Weigel, Young, Cleveland & Dekkers, 2014). En bovinos productores de carne SG no ha sido utilizada tan ampliamente como en bovinos lecheros, en parte por su menor exactitud y porque las ventajas económicas no son tan grandes (Meuwissen et al., 2016).

Para varias características importantes, difíciles de medir como conversión alimenticia y calidad de la canal, por el alto costo que implicaría establecer poblaciones de entrenamiento, una alternativa sería posible al establecer una población de entrenamiento multirracial y utilizar análisis no lineales basados en paneles de alta densidad (Meuwissen et al., 2016).

Los programas de mejoramiento genético en porcinos no siempre cuentan con los menores intervalos entre generaciones cuando se utiliza el comportamiento productivo de los animales cruzados en la predicción de VGs de los candidatos a selección. En estos casos la selección asistida por marcadores utilizando el

genoma completo, tiene el potencial de aumentar la ganancia genética al identificar con precisión a los mejores candidatos de selección a una edad temprana, disminuyéndose así el intervalo entre generaciones.

Evaluaciones genómicas utilizando genotipos densos han demostrado aumentos en la exactitud de los VGn en cerdos (Cleveland, Hickey & Forni, 2012). Sin embargo, la mejoría se ve limitada por el gran número de progenie en cada camada que nace, lo que hace que el costo de la genotipificación de todos los candidatos se vuelva prohibitivo, por lo cual se han desarrollado métodos para imputar genotipos densos utilizando paneles de SNP de baja densidad (Hickey et al., 2012; VanRaden, O'Connell, Wiggans & Weigel, 2011).

La implementación de SG en porcinos está mayormente dirigida a características cuya medición es invasiva como calidad de la canal, características maternas que no pueden medirse en machos y características de comportamiento productivo en animales cruzados, que no pueden determinarse en animales puros (Meuwissen et al., 2016).

2.1.5 Programas de mejoramiento genético con selección genómica

En la biología molecular moderna se han adoptado dos enfoques básicos para detectar QTL: cartografía de QTL para regiones cromosómicas usando marcadores genéticos y el estudio de la variación en genes candidatos seleccionados con base en su papel fisiológico conocido. Ambos métodos han dado lugar a descubrimientos importantes, pero ha resultado difícil de descubrir los suficientes QTL para explicar la alta proporción de la variación genética en muchas características. Poco a poco se ha dejado de estudiar la detección de QTL para dar paso a las evaluaciones genómicas de los animales, área aparentemente más promisorio (Dekkers, 2004).

La SG se refiere al uso de marcadores distribuidos en todo el genoma, que permiten predecir el valor genético para una característica de los candidatos sujetos a selección. Para predecir dicho valor se requiere una distribución previa de los efectos del marcador, si el marcador se distribuye normalmente con una varianza constante, el BLUP se puede utilizar para predecir los efectos estimados de los marcadores y posteriormente los VGn. Se supone que los marcadores son lo bastante densos y que se encuentran en desequilibrio de ligamiento con el QTL. Por tanto, los marcadores pueden utilizarse para seleccionar el alelo favorable en cada QTL, sin identificar el QTL o el polimorfismo funcional (Brøndum et al., 2011). Mediante simulación se ha demostrado que el VGn de un individuo con sólo datos de marcadores se puede obtener con una precisión de hasta 0.85 con una PEn con información de marcadores y fenotipos (Daetwyler, Swan, van der Werf & Hayes, 2012a; Meuwissen et al., 2001).

Los primeros años de investigación en genómica, hasta aproximadamente 2005, produjeron poco en términos de desarrollos tecnológicos aplicables a programa de mejoramiento genético. Actualmente, en todas las especies de interés zootécnico se han implementado programas de SG. Se espera que estos programas se consoliden en el futuro (Ibanez-Escriche & Simianer, 2015). Esta área requiere investigación para resolver las particularidades que cada especie tiene (Meuwissen et al., 2016).

2.2 Desequilibrio de ligamiento

Dos genes ligados se segregan juntos, si ambos se encuentran en el mismo cromosoma a una distancia menor que 50% de la fracción de recombinación (50 cM ó 120 Mpb). Genes localizados en cromosomas diferentes o en el mismo cromosoma pero distanciados a más de 50 cM segregan independientemente durante la meiosis dando lugar a gametos con diferentes alelos. Por tanto, se espera mayor probabilidad de segregación independiente en dos genes localizados en diferentes cromosomas (Van Vleck, Pollak & Oltenacu, 1987).

Solberg, Sonesson, Woolliams & Meuwissen (2008) mencionan que el desequilibrio de ligamiento (DL) es el factor clave que impulsa el proceso de predicción genómica y se refiere a la asociación no aleatoria o independiente de alelos de diferentes *loci*, esto es, la probabilidad de encontrar un alelo de un *locus* en una población dependiente de la frecuencia del alelo de otro *locus*; por lo que la frecuencia de haplotipos de dos o más alelos de distintos *loci* será mayor de lo esperado con relación a haplotipos independientes.

El desequilibrio de ligamiento no implica que dos *loci* estén ligados, por lo que dichos *loci* podrían estar en el mismo o distinto cromosoma. Sin embargo, dos *loci* fuertemente ligados permite identificar un desequilibrio de ligamiento y un mapeo asociativo más confiable para estimadores de QTL (Brøndum et al., 2011).

Un marcador polimórfico informativo permite rastrear a través de generaciones, un fragmento particular del genoma. Para que un marcador sea útil necesita estar en fuerte DL con el gen causal, entre mayor sea el DL, mejor su utilidad. Si un marcador tiene una frecuencia alélica muy diferente de la del gen causal, dado que las frecuencias alélicas cambian y suelen fijarse las de mayor proporción bajo selección artificial a través de generaciones; idealmente se requiere de un amplio rango de marcadores aledaños que difieran en frecuencia alélica (Wientjes, Veerkamp & Calus, 2012).

La secuenciación del genoma bovino y el desarrollo de tecnologías que permiten la caracterización simultánea de decenas de miles de marcadores genéticos han abierto oportunidades para explotar el DL, más que tener que crearlo, ya que es un fenómeno que generalmente ocurre en animales y plantas. La extensión del DL se incrementa en promedio si el marcador está físicamente cerca del gen causal desconocido (Segelke et al., 2012).

2.2.1 Causas de desequilibrio de ligamiento

El desequilibrio de ligamiento se genera de varias formas, una de ellas es mediante selección, puesto que al seleccionar se alteran las frecuencias de los gametos favoreciendo a unas frente a otras, con esto la independencia estadística se pierde (Erbe et al., 2012). Cuando hay DL, por ejemplo al cruzar dos razas, existen caracteres que aparecen como más relacionados de lo que están por la conjunción génica entre *loci* no ocurrentes en alguna de las razas (transposición), pero su asociación se va perdiendo con el tiempo. Debe tenerse cuidado al estimar correlaciones genéticas en poblaciones en las que hay DL, por ejemplo en aquellas que han sido seleccionadas (Calus, Huang, Vereijken, Visscher, Napel & Windig, 2014).

2.2.2 Implicaciones de desequilibrio de ligamiento

Las asociaciones entre marcadores moleculares (SNPs, microsatélites, etc.) y genes (tanto QTLs como genes menores) desaparecen conforme avanzan las generaciones. Esto implica que no se puedan aplicar las asociaciones encontradas en una generación a datos antiguos, ni a datos nuevos después de algunas generaciones (Bolormaa et al., 2013).

Puede haber marcadores no ligados a genes cuantitativos que estén en DL con ellos, de manera que seleccionando el marcador se seleccione también el gen de interés, aunque estas asociaciones desaparezcan rápidamente en las siguientes generaciones por cambios en sus frecuencias alélicas. Su utilización se basa simplemente en que no estén en estado de independencia estadística (Larmer, Sargolzaei & Schenkel, 2014).

Con el modelo infinitesimal se asume que los caracteres siguen una distribución normal porque están controlados por muchos genes independientes con un pequeño efecto cada uno de ellos. La selección produce DL, dependencia

estadística, por lo que no se cumple una de las condiciones del teorema del límite central y no se puede invocar la normalidad en este caso. Aunque existe la sensación de que este efecto no es importante, depende de diferentes circunstancias, como tamaño efectivo de la población o intensidad de la selección (Erbe et al., 2012).

Las escalas de tiempo que reducen la vinculación con DL son muy largas, con la participación de muchas generaciones. Un ejemplo son los muy cortos tramos del genoma que las diferentes razas pueden compartir, como resultado de antepasados comunes anteriores a la divergencia de la raza. En lugar de concentrarse en DL entre SNP y QTL, la predicción genómica puede utilizar tanto la vinculación como el DL (Daetwyler, Kemper, van der Werf & Hayes, 2012b).

Otra consecuencia es que la varianza aditiva total ya no es la suma de la de cada gen, sino que hay que considerar las covarianzas entre genes; esto es particularmente importante en el caso de que la población esté seleccionada, porque selección produce DL y estas covarianzas ya no son nulas. De hecho son negativas, aunque parezca contra-intuitivo (da la impresión de que al seleccionar, los genes que van en la misma dirección deberían aumentar todos su frecuencia y las covarianzas ser positivas), y la varianza genética aditiva disminuye (Erbe et al., 2012).

2.3 Exactitud de valores genómicos

La precisión en la estimación de valores genéticos-genómicos es un parámetro importante por lo siguiente: la respuesta a la selección es proporcional a la exactitud, por lo que la exactitud refleja la ganancia genética potencial de un esquema de mejoramiento genético; y la exactitud refleja la credibilidad del VG individual y se relaciona con el riesgo de que este cambie a través de generaciones sucesivas y cuando se disponga de más información fenotípica y

molecular (Bijma, 2012).

La predicción genómica utiliza marcadores SNP para predecir los valores genéticos de los candidatos. Por tanto, la mejor variable de respuesta para estimar los efectos SNP es el verdadero valor genético (Guo et al., 2010). En la actualidad se han generado grandes cantidades de información genómica para casi todas las especies de interés zootécnico y su uso ha dado lugar a aumentos en la exactitud de los VGn (VanRaden et al., 2009).

Una alta precisión de los valores genéticos predichos es crucial para lograr un alto progreso genético. Establecer correctamente la PEn es un paso importante en un programa de selección animal en el que se lleva a cabo la evaluación genómica, ya que una PEn diseñada de manera óptima permite maximizar la precisión para la población dada (Pszczola, Strabel, Mulder & Calus, 2012).

La exactitud de valores obtenidos en SG depende del número, la distribución y las contribuciones de los genotipos y fenotipos de la evaluación genómica. Cuando la PEn tiene sólo animales de un sexo, el mayor beneficio es para los animales jóvenes con genotipo del mismo sexo; sin embargo, cuando son genotipificados ambos sexos, la cantidad de información genómica aumenta en gran medida, así como la exactitud de los VGn (Lourenco et al., 2015).

Los aumentos de precisión se deben principalmente a la mejora en las inferencias en las relaciones entre individuos y el DL entre QTL y marcadores moleculares. Los dos componentes están probablemente correlacionados, porque una relación cercana entre subgrupos y dentro de ellos incrementa el DL (Daetwyler et al., 2012b). La medida en que las dos fuentes contribuyen a la exactitud de la predicción es muy clara en la actualidad. Sin embargo, la distinción entre ellas es importante ya que la exactitud debido a DL es más probable que persista a través de generaciones y razas, que la exactitud debido a las relaciones entre individuos

(De Roos, Hayes & Goddard, 2009; Habier, Fernando & Dekkers, 2007; Meuwissen et al., 2001).

Exactitudes mayores en los valores genómicos se obtienen cuando la relación entre animales en la PEn es débil y la relación entre las poblaciones de entrenamiento y la de evaluación es alta (Alarcón-Zúñiga et al., 2015; Pszczola et al., 2012). La predicción genómica entre individuos estrechamente relacionados se basa en la vinculación mientras que la predicción de individuos alejados por otro lado requiere DL entre QTL y marcadores (Daetwyler et al., 2012a).

La inclusión de la información de los marcadores en la selección de los jóvenes candidatos a la selección, en promedio mejora la exactitud en 0.04 en comparación con la selección por prueba de progenie. Para evaluar la exactitud de las predicciones genómicas, es importante reducir la dependencia entre PEn y PEv, importante para la estimación de las exactitudes genómicas (Thomassen, Guldbrandtsen, Su, Brøndum & Lund, 2012).

Para mantener el nivel inicial de exactitud la PEn necesita actualizarse cada determinado tiempo, esto es, volver a ajustar los valores genómicos a través del genotipado de SNP en los individuos que conforman la generación actual. No sólo el número de los animales que comprenden esta población es importante, sino también la selección de los mismos; de esta manera el impacto en la precisión a través de diferentes actualizaciones de la PEn es grande (Pszczola et al., 2012).

2.3.1 Factores que afectan la exactitud de los valores genómicos

Uno de los factores más importantes que influyen en la exactitud de los VGn, es el número de registros fenotípicos en PEn (Brito, Braccini, Sargolzaei, Cobuci &

Schenkel, 2011). Saatchi, Miraei-Ashtiani, Nejati-Javaremi, Moradi-Shahrehabak & Mehrabani-Yeghaneh (2010) encontraron que el aumento en la cantidad de registros fenotípicos, incluso para individuos de generaciones más viejas, incrementó las exactitudes de los VGn, por ejemplo cuando se incluyeron individuos de las generaciones 51 y 52 a la PEn que ya incluía las generaciones 53 a 58, el aumento en la exactitud fue de 0.737 a 0.749. El número de registros fenotípicos en la PEn tuvo un importante papel en la exactitud de los VGn, especialmente en características con bajo índice de herencia.

Otro factor importante que afecta la exactitud de la predicción genómica es la densidad de marcadores (Solberg et al., 2008; Meuwissen, 2009; Weigel et al., 2009; Habier, Tetens, Seefried, Lichtner & Thaller, 2010). Una densidad de marcadores más alta significa que, en promedio, ellos están en un desequilibrio de enlace más fuerte con genes que afectan al rasgo de interés, lo que debería conducir a mejores predicciones genómicas (Su et al., 2012).

La distancia entre PEn y PEv es también un factor importante que influye sobre la exactitud de valores genómicos predichos (Alarcón-Zúñiga et al., 2015). Saatchi et al. (2010) encontraron mayores exactitudes, 0.634 vs 0.555, cuando se utilizaron individuos de generaciones recientes en PEn que cuando se utilizaron individuos de generaciones más distantes. Aún con un número menor de registros fenotípicos de individuos de generaciones recientes obtuvieron exactitudes de los VGn mayores. Por ejemplo, con 2,000 registros de las generaciones 55 a la 58 en PEn se obtuvo una exactitud de 0.706, en comparación con 0.695 obtenida con 3,000 registros de las generaciones 51 a 56.

Meuwissen, Hayes & Goddard (2001) estudiaron el efecto de tres diferentes números de registros fenotípicos en la PEn en la exactitud de los VGn en la población de validación, mediante el uso de tres diferentes métodos estadísticos

y encontraron que en todos ellos la exactitud aumentó al incrementar el número de registros en PEn. Similares resultados se reportaron en otros estudios (Brito et al., 2011; Calus & Veerkamp, 2007; Muir, 2007).

El número de marcadores moleculares influye la exactitud de los VGn. Brito et al. (2011) encontraron que las exactitudes de valores genómicos directos se incrementaron significativamente ($P < 0.05$) cuando se incrementó el número de marcadores moleculares de 40 a 80 mil. El número de *loci* causales también influye en la exactitud de los valores genómicos predichos. Kizilkaya, Fernando & Garrick (2010) demostraron que la exactitud a través de razas utilizando 1,000 animales en PEn fue solamente de 0.2 a 0.4, cuando el número de *loci* causales varió de 50 a 500. Esto adicionalmente demuestra que PEn y la PEv preferentemente deben ser de la misma raza.

Cuando se utiliza información *derregresada*, ajustada por la contribución de los padres, para la predicción de VGn, además de obtenerse ventajas hasta 2.76 veces más que un registro sencillo de un individuo (Garrick, Taylor & Fernando, 2009), también permite obtener exactitudes mayores. Por ejemplo, Ostersen et al. (2011) estimaron exactitudes de 18 a 39% mayores de los VGn cuando utilizaron valores genéticos *derregresados* en lugar de los VG predichos, como variable de respuesta. La estructura poblacional también influye en la exactitud de la predicción de los VGn (Goddard & Hayes, 2009; Habier et al. 2010).

Kachman et al. (2013) utilizaron poblaciones de entrenamiento de una raza y de evaluación de otra. Estos autores encontraron que aún en razas relativamente cercanas, se observó limitada exactitud en la predicción de VGn cuando la raza evaluada no estuvo representada en PEn. Sin embargo, adicionar información de otras razas a una población de entrenamiento de una raza, posibilita obtener VGn más robustos a través de razas.

Por su parte, Habier et al. (2010) demostraron que las exactitudes de los VGn dependen tanto de las relaciones genéticas entre individuos en PEn y PEv, como del DL entre marcadores y QTL. Otros factores que influyen en la exactitud de los VGn son la magnitud de la heredabilidad de la característica y el número de marcadores usados para la predicción genómica (Meuwissen et al., 2001; Hayes et al., 2009a). En bovinos Holstein, Aguilar et al. (2010) encontraron que el uso de frecuencias iguales de los alelos, resultó en exactitudes mayores y sesgos menores.

2.4 Simulación de datos en selección genómica

La simulación es una herramienta valiosa para evaluar y validar los métodos propuestos para la cartografía QTL y la selección genómica a un costo muy bajo, permitiendo también la predicción de futuros cambios en los parámetros genéticos. Durante las últimas décadas, la simulación ha jugado un papel importante en genética de poblaciones y selección genómica (Sargolzaei & Schenkel, 2009).

Dentro de los campos de la genética que participan en el estudio y la aplicación de la selección genómica (sobre todo en mejoramiento genético de animales y plantas), el desarrollo de métodos para simular correctamente los datos requiere un mayor enfoque de la investigación (Daetwyler, Calus, Pong-Wong, De los Campos & Hickey, 2013).

Datos genómicos reales y simulados se han utilizado en estudios de predicción genómica para investigar diversos aspectos tales como la eficacia de los diferentes métodos de análisis, comparación de programas alternativos de mejoramiento genómico y la exploración de la dinámica de la selección genómica de corto, mediano y largo plazo. La simulación de bases de datos para mejorar el método de selección genómica se realiza tanto en animales (Daetwyler et al., 2013; Gupta, Conrad, Spötter, Reinsch & Bienefeld, 2012) como en plantas (Hickey et al., 2014).

2.4.1 Implicaciones

Una población base se utiliza como punto de partida para los estudios de simulación, esta población se compone de individuos en un pedigrí para el cual no hay información ancestral disponible, y se supone que está en desequilibrio de ligamiento (Gupta et al., 2012).

Los datos reales ofrecen la ventaja de reflejar la complejidad, mientras que los datos simulados permiten explorar aspectos importantes como la arquitectura genética del rasgo, el número de marcadores utilizados para el análisis y el grado de relación entre PEn y PEv. La simulación es útil porque permite pruebas rápidas y repetidas de una amplia gama de hipótesis a bajo costo; por ejemplo, la viabilidad inicial de la selección genómica o el impacto del tamaño de la población de referencia. Sin embargo, la simulación de genomas y mecanismos causales en el ganado es compleja. Hay muchas formas diferentes de variabilidad genómica y una amplia variedad de historias de poblaciones plausibles, así como una considerable incertidumbre sobre cómo varían las tasas de mutación y recombinación, y sobre el tipo y distribución estadística de la acción génica (Daetwyler et al., 2013).

Los estudios de simulación proporcionan información sobre algunos mecanismos que influyen la exactitud del VG, pero se basan en supuestos para la generación de datos, cuya validez se desconoce en datos reales, como número de QTL, distribución de sus efectos y la población histórica que contribuyó a la actual estructura de DL (Wolc et al., 2011).

2.4.2 Hallazgos

Los genotipos y los resultados genómicos están ampliando rápidamente los datos y las herramientas disponibles para los criadores. La densidad de marcadores muy alta aumenta la exactitud de la selección dentro de la raza ligeramente

(1.6%) en la simulación, mientras que las densidades más bajas permiten a los criadores aplicar SG rentable a muchos más animales (VanRaden et al., 2011).

Wientjes et al. (2012) encontraron que el tamaño de la población de entrenamiento influye en el efecto del DL y de las relaciones familiares en la exactitud de la predicción genómica; PEv pequeñas dan mayor efecto de las relaciones familiares en comparación con el DL, y mayores poblaciones de referencia en un mayor efecto de DL en la exactitud. Por su parte, y de manera similar, Brito et al. (2011) estudiaron el efecto del tamaño de PEn y el grado de DL en la exactitud de VGn. Los resultados de este estudio de simulación, mostraron que mientras más grande sea PEn y los paneles de marcadores más densos, con el objetivo de aumentar el DL entre los marcadores y QTL, la exactitud de los VGn predichos aumentará.

En acuicultura, Nielsen, Sonesson & Meuwissen (2011) estudiaron la ventaja en términos de ganancia genética de la aplicación de selección genómica en un programa de cría acuícola y obtuvieron un incremento de 81%, comparado con un esquema de cría tradicional de peces.

En general, los estudios de simulación se realizan para reafirmar el hecho de requerirse un mayor número de animales con genotipos y fenotipos para lograr una alta precisión en la selección de estos rasgos. Entonces, invertir en genotipificación de animales con registros es especialmente beneficioso (Lillehammer, Meuwissen & Sonesson, 2011).

2.5 Literatura citada

Aguilar, I., I. Misztal, A. Legarra & S. Tsuruta. (2010). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *Journal of Animal Breeding and Genetics*, 128, 422-428.

- Alarcón-Zúñiga, B., F. Ramírez-Flores, A. Ruíz-Flores, R. Ramírez-Valverde, L. A. Saavedra-Jiménez & J. L. Zepeda-Batista. (2015). Comparación de la exactitud de valores genómicos de animales predichos a través del análisis con dos modelos alternativos. *Agrociencia*, 49, 613-622.
- Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics*, 2, 130-138.
- Bijma, P. (2012). Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *Journal of Animal Breeding and Genetics*, 129, 345-358.
- Bolormaa, S., J. E. Pryce, K. Kemper, K. Savin, B. J. Hayes, W. Barendse, ... & M. E. Goddard. (2013). Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. *Journal of Animal Science*, 91, 3088-3104.
- Brito, F. V., J. Braccini N., M. Sargolzaei, J. A Cobuci & F. S Schenkel. (2011). Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics*, 12, 80-89.
- Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbbrandtsen, W. F. Fikse & M. S. Lund. (2011). Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *Journal of Dairy Science*, 94, 4700-4707.
- Calus, M. P. L. & R. F. Veerkamp. (2007). Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics*, 124, 362-368.

- Calus, M. P. L., H. Huang, A. Vereijken, J. Visscher, J. ten Napel & J. J. Windig. (2014). Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genetics Selection Evolution*, 46, 1-57.
- Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra & W. M. Muir. (2011). Effect of different genomic relationship matrices on accuracy and scale. *Journal of Animal Science*, 89, 2673-2679.
- Christensen, O. F. & M. S. Lund. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42, 1-8.
- Cleveland, M. A., J. M. Hickey & S. Forni. (2012). A common dataset for genomic analysis of livestock populations. *G3: Genes, Genomes, Genetics*, 2, 429-435.
- Cleveland, M. A. & J. M. Hickey. (2013). Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. *Journal of Animal Science*, 91, 3583-3592.
- Daetwyler, H. D., A. A. Swan, J. H. J. van der Werf & B. J. Hayes. (2012a). Accuracy of pedigree and genomic predictions of carcass and novel quality traits in multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution*, 44, 1-33.
- Daetwyler, H. D., K. E. Kemper, J. H. J. van der Werf & B. J. Hayes. (2012b). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal of Animal Science*, 90, 3375-3384.
- Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos & J. M. Hickey. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting and benchmarking. *Genetics*, 193, 347-365.

- De los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler & M. P. L. Calus. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193, 327-345.
- De Roos, A. P. W., B. J. Hayes & M. E. Goddard. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, 183, 1545-1553.
- Dekkers, J. C. M. (2004). Commercial application of marker and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science*, 82, 313-328.
- Dekkers, J. C. M. & F. Hospital. (2002). The use of molecular genetics in improvement of agricultural populations. *Nature Review Genetics*, 3, 22-32.
- Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, ... & M. E. Goddard. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95, 4114-4129.
- Fernando, R. L., & M. Grossman. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, 21, 467-477.
- Garrick, D. J. (2011). The nature, scope and impact of genomic prediction in beef cattle in the United States. *Genetics Selection Evolution*, 43, 17-27.
- Garrick, D. J. & M. Saatchi. (2011). Opportunities and challenges for genomic selection of beef cattle. *Revista Brasileira de Zootecnia*, 40, 310-316.
- Garrick, D. J., J. F. Taylor & R. Fernando. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41, 55-62.
- Goddard, M. E. & B. J. Hayes. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124, 323-330.

- Goddard, M. E. & B. J. Hayes. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10, 381-391.
- Goddard, M., B. Hayes & T. H. E. Meuwissen. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics*, 128, 409-421.
- Guo, G., M. S. Lund, Y. Zhang & G. Su. (2010). Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *Journal of Animal Breeding and Genetics*, 127, 423-432.
- Gupta, P., T. Conrad, A. Spötter, N. Reinsch & K. Bienefeld. (2012). Simulating a base population in honey bee for molecular genetic studies. *Genetics Selection Evolution*, 44, 1-14.
- Habier, D., R. L. Fernando & J. C. M. Dekkers. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177, 2389-2397.
- Habier, D., R. L. Fernando & J. C. M. Dekkers. (2009). Genomic selection using low-density marker panels. *Genetics*, 182, 343-353.
- Habier, D., J. Tetens, F. R. Seefried, P. Lichtner & G. Thaller. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, 42, 1-5.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain & M. E. Goddard. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92, 433-443.

- Hickey, J. M., J. Crossa, R. Babu & G. de los Campos. (2012). Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science*, 52, 654-663.
- Hickey, J. M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B. M. Prasanna, ... & G. Gorjanc. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, 54, 1476-1488.
- Ibanez-Escriche, N. & H. Simianer. (2015). Animal breeding in the genomics era. *Animal Frontiers*, 5, 4-5.
- Kachman, S. D., M. L Spangler, G. L. Bennett, K. J. Hanford, L. A. Kuehn, W. M. Snelling, ... & E. J. Pollak. (2013). Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution*, 45, 30-38.
- Kizilkaya, K., R. L. Fernando & D. J. Garrick. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *Journal of Animal Science*, 88, 544-551.
- Larmer, S. G., M. Sargolzaei & F. S. Schenkel. (2014). Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *Journal of Dairy Science*, 97, 3128-3141.
- Legarra, A., I. Aguilar & I. Misztal. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92, 4656-4663.
- Lillehammer, M., T. H. E. Meuwissen & A. K. Sonesson. (2011). Genomic selection for maternal traits in pigs. *Journal of Animal Science*, 89, 3908-3916.

- Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, ... & I. Misztal. (2015). Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. *Genetics Selection Evolution*, 47, 1-56.
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution*, 41, 1-35.
- Meuwissen, T. H. E., B. J. Hayes & M. E. Goddard. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.
- Meuwissen, T., B. Hayes & M. Goddard. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6, 6-14.
- 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, 342-355.
- Nielsen, H. M., A. K. Sonesson & T. H. E. Meuwissen. (2011). Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *Journal of Animal Science*, 89, 630-638.
- Nieuwhof, G.J., K.T. Beard, K.V. Konstantinov, P.J. Bowman & B.J. Hayes. (2010). Implementation of Genomics in Australia. *Interbull Bulletin*, 42, 35-39.
- Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su & P. Madsen. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution*, 43, 38-43.

- Pszczola, M., T. Strabel, H. A. Mulder & M. P. L. Calus. (2012). Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science*, 95, 389-400.
- Saatchi, M., S. R. Miraei-Ashtiani, A. Nejati Javaremi, M. Moradi-Shahrehabak & H. Mehrabani-Yeghaneh. (2010). The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *African Journal of Biotechnology*, 9, 438-442.
- Sargolzaei, M. & F. S. Schenkel. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25, 680-681.
- Sargolzaei, M., J. P. Chesnais & F. S. Schenkel. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15, 478.
- Schenkel, F. (2012). Inbreeding using genomics and how it can help. World Holstein Federation Conference, Toronto.
http://www.whff.info/documentation/documents/2012_03_FlavioSchenkel.pdf
- Segelke, D., J. Chen, Z. Liu, F. Reinhardt, G. Thaller & R. Reents. (2012). Reliability of genomic prediction for German Holsteins using imputed genotypes from low-density chips. *Journal of Dairy Science*, 95, 5403-5411.
- Solberg, T. R., A. K. Sonesson, J. A. Woolliams & T. H. E. Meuwissen. (2008). Genomic selection using different marker types and densities. *Journal of Animal Science*, 86, 2447-2454.
- Soller, M. (1978). The use of loci associated with quantitative effects in dairy cattle improvement. *Animal Production*, 27, 133-139.

- Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand & M. S. Lund. (2012). Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *Journal of Dairy Science*, 95, 4657-4665.
- Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum & M. S. Lund. (2012). Reliabilities of genomic estimated breeding values in Danish Jersey. *Animal*, 6:5, 789-796.
- Van Eenennaam, A.L., K.A. Weigel, A.E. Young, M.A. Cleveland & J.C. Dekkers. (2014). Applied animal genomics: Results from the field. *Annual Review Animal Biosciences*, 2, 105-139.
- Van Vleck, L. D., E. J. Pollak & E. A. B. Oltenacu. (1987). *Genetics for the Animal Sciences*. W. H. Freeman. New York, NY, USA. 391 p.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor & F. S. Schenkel. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24.
- VanRaden, P., J. R. O'Connell, G. R. Wiggans & K. A. Weigel. (2011). Genomic evaluations with many more genotypes. *Genetics Selection Evolution*, 43, 1-10.
- Weber, K. L., D. J. Drake, J. F. Taylor, D. J. Garrick, L. A. Kuehn, R. M. Thallman, ... & A. L. Van Eenennaam. (2012). The accuracies of DNA-based estimates of genetic merit derived from Angus or multibreed beef cattle training populations. *Journal of Animal Science*, 90, 4191-4202.
- Weigel, K. A., G. de los Campos, O. González-Recio, H. Naya, X. L. Wu, N. Long, ... & D. Gianola. (2009). Predictive ability of direct genomic values for lifetime

net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *Journal of Dairy Science*, 92, 5248-5257.

Wientjes, Y. C. J., R. F. Veerkamp & M. P. L. Calus. (2012). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics*, 112, 1-37.

Wolc, A., C. Stricker, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, ... & J. C. M. Dekkers. (2011). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution*, 43, 1-23.

3 ACCURACY OF GENOMIC VALUES PREDICTED USING DEREGRESSED PREDICTED BREEDING VALUES AS RESPONSE VARIABLES

3.1 Abstract

Highly accurate predicted genetic values must be obtained at an early age to promote rapid genetic progress. The objectives of this study were to compare accuracies (R^2) of genomic values (GVs) and to estimate genetic correlation between true genetic values and genomic values obtained using predicted breeding values (EBV) and deregressed EBV (DEBV) as response variables. A first population, effective population size 800 and 100 generations, was simulated using the QMSim program to generate linkage disequilibrium. Thereafter, 20 males and 200 females were used to generate a second 14-generation population, with 6,400 individuals per generation and its corresponding phenotype and genotype in SNP terms. Generations 7 to 14 of the second population were used in several combinations as training (PEn) and evaluation (PEv) subpopulations. GV, their accuracies, and genetic correlations were obtained using the GenSel and ASREML programs. When PEn was the largest, the mean R^2 of GV was the highest, 0.77 ± 0.01 . The closer PEn was to PEv, the higher the R^2 , and correspondingly, the lower the predicted error variance. The trends for R^2 and PEV held true for both EBV and DEBV used as response variables. Genetic correlation estimates between true genetic values and GVs varied from 0.41 to 0.53 in the two scenarios studied. They decreased when PEn and PEv were farther apart. There were only slight advantages of using DEBVs as response variables over using EBVs.

Key words: Genomic evaluation, deregressed predicted genetic value, genomic predicted value, accuracy, genetic correlation.

3.2 Introduction

Genetic improvement depends on genetic variation, selection intensity, generation interval, and accuracy of estimated breeding values (EBV). In the genetic evaluation of animals, it is important to maximize EBV accuracy. An increase in EBV accuracy for selection of candidate animals will spur genetic progress. Among other advantages, the use of genomic selection allows an increase in the accuracy of genetic values (Goddard, 2009), especially at a young age (Knol et al., 2016). Three technological breakthroughs have boosted widespread DNA information use in animal breeding (Meuwissen et al., 2016): the development of genomic selection technology, the discovery of massive numbers of genetic markers (SNPs), and high throughput cost-effective genotyping technology. Although the advantages of genomic selection have been observed most notoriously in dairy cattle (Ibanez-Escriche & Simianer, 2015), in general, the use of genomic selection can be expected to yield improvements in genetic progress of up to 10% in any species (Ibanez-Escriche & Simianer, 2015).

In genomic evaluation, response variables can be individual phenotypes, repeated observations, records on close family members such as progeny, EBVs or their deregressed counterparts from genetic evaluations (Guo et al., 2010; Ibanez-Escriche & Simianer, 2015). According to these authors, using deregressed EBV (DEBV), an accuracy of up to 2.76 times higher than with records of a single individual can be obtained. With average daily gain and feed conversion ratio of swine data, obtained accuracies were 18 to 39% higher, depending on the trait evaluated, when DEBVs were used as response variables instead of EBVs (Ostersen et al., 2011). These authors concluded that DEBV is the preferred response variable, whereas the choice of statistical method was less critical when they analyzed purebred swine data. The increase of 18 to 39% in reliability is worthwhile, since the reliabilities of the genomic breeding values directly affect the returns from genomic selection (Ostersen et al., 2011).

Deregressed EBVs, with the parent average removed, produce more exact predicted genomic values (GV) for two reasons (Garrick et al., 2009). First,

DEBVs, when used as the response variable, result in fewer double counts than when EBVs are used because the DEBVs exclude information from the individual's ancestors. If both the offspring and its parents are genotyped, the degree of double count decreases when DEBVs are used as the response variable. Second, when using EBVs as the response variable, the degree of double count in the GVs decreases, particularly when the reliabilities of the genetic values are low.

However, DEBVs are not always the best choice for use as the response variable in genomic evaluation. Simulated dairy cattle (Guo et al., 2010) and jumping horse (Ricard et al., 2013) data were used to compare EBVs and DEBVs as response variables. Both groups of authors found only slight advantages of DEBVs, over conventional EBVs, as response variables. The objectives of this study were to compare the R^2 of GVs and to estimate the genetic correlation between true genetic values and GVs obtained using predicted breeding values (EBV) and deregressed EBV (DEBV) as response variables for four training populations and four evaluation generations.

3.3 Material and Methods

3.3.1 Data simulation

The methodology for simulating the training (PEn) and evaluation (PEv) populations used in this study was described previously (Alarcón-Zúñiga et al., 2015). Briefly, two populations using the QMSim program (Sargolzaei & Schenkel, 2009) were simulated considering the characteristics of dairy cattle. The first, to obtain linkage disequilibrium, had 800 individuals as the effective population size and 100 generations. The second population, where PEn and PEv originated, had 14 discrete generations, each of which was generated randomly using 20 males and 200 females, a panel of 53,010 SNPs (each evenly separated by 100 centiMorgans) randomly placed in 30 chromosomes, and 540 QTLs with effects coming from a gamma distribution (Weller et al., 2005). Both SNPs and QTLs were regarded as biallelic with random starting frequencies. Genotypes and

phenotypes of 6,400 individuals were simulated; the heritability used was 0.4 and only additive effects were considered. Genotypic and phenotypic information was generated using the QMSim program (Sargolzaei & Schenkel, 2009). The four PEn comprised generations 10 (n=1000); 9 and 10 (n=1400); 8 to 10 (n=1800); and 7 to 10 (n=2200); as well as their phenotypes and the corresponding EBVs and DEBVs. The four PEv comprised generations 11 to 14.

3.3.2 Response variables

In a first step, the EBVs were predicted with a single-trait animal model including the random effect of animal, the fixed effects of sex and generation of the individual. The ASREML program (Gilmour et al., 2009) was used at this stage. The DEBVs were then obtained following Garrick, Taylor and Fernando (2009). Weight (w_i) for the i^{th} animal was obtained using the following equation (Garrick et al., 2009):

$$w_i = (1 - h^2) / [(c + (1 - r^2) / r^2) h^2],$$

where c is the lack of fit of the prediction equation, or the genetic part not explained by the markers (Garrick et al., 2009); the value assumed was $c=0.1$; heritability of the trait, h^2 , was assumed to be 0.4; and r^2 was the reliability of the DEVGs for the i^{th} animal.

Deregression of EBVs adjusts for ancestral information, it removes shrinkage present in EBV, and by taking parental contribution into account, DEBVs can be regarded as equivalent to the information provided by the records of each sire and its progeny (Weber et al., 2012).

3.3.3 Genomic analysis

In a second step, the predicted breeding values (EBVs) obtained using ASREML and their corresponding DEBVs were used as response variables to predict the GVs. A weighted genomic analysis was carried out using the BayesC π function of the Gen-Sel program (Fernando & Garrick, 2009). A 41,000-round long chain

was used. The last 40,000 samples were used to obtain the *a posteriori* mean estimates of marker effects and variances. The first 1,000 iterations were regarded as the burn-in period; α was fixed at 0.95. The genomic analysis used animals of generations 7 to 10 to obtain the prediction equations. The evaluation populations were generations 11 to 14. The Bioinformatics to Implement Genomic Selection (BIGS) platform (<http://biggs.ansci.iastate.edu/>) was used for the analysis.

The genomic values and their corresponding accuracies were obtained by summing all the SNP effects, using the following equation:

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

where GV_{ni} is the genomic value for the i^{th} individual, z_{ij} is the genotype of the j^{th} marker on the i^{th} individual, and $\hat{u}_j$ is the *a posteriori* mean of SNP effect for the j^{th} marker.

Accuracies (R^2) of GVs were obtained as the square of the correlation between GVs and the true genetic values (Guo et al., 2010; Pszczola et al., 2012a; Weber et al., 2012). Criteria for comparing the two alternatives of analysis were R^2 and GV prediction error variance (PEV). Additionally, as another criterion for comparing the two response variables studied, we used the genetic correlation estimates between the true genetic values and the predicted GVs from the two alternatives of genomic analysis (Guo et al., 2010). These estimates were obtained using ASREML (Gilmour et al., 2009).

3.4 Results and Discussion

3.4.1 Accuracy of genomic values

An important aspect in genetic improvement is the response to selection, and this depends on selection accuracy (Pszczola et al., 2012b). Table 1 shows the means and corresponding standard deviation for R^2 and PEV of GVs obtained from the

different combinations of PEn and PEv when the response variable was DEBV. The highest R^2 , 0.77 ± 0.01 , was observed for the largest training population (generations 7 to 10, $n=2200$ individuals) and 11 was the generation under evaluation. In contrast, the lowest mean for R^2 , 0.28 ± 0.06 , was observed for the combination of the smallest training population and the farthest evaluation population being evaluated, generation 14. These results are within the range of R^2 values reported by Hassani, Saatchi, Fernando and Garrick (2015), who found $0.49 (\pm 1 \text{ SNP})$ to $0.75 (\pm 100 \text{ SNPs})$ using whole-genome training for single QTL with a 50 K SNP panel and BayesC0.

Two clear trends can be observed for R^2 in Table 1. First, as PEn and PEv moved farther apart, R^2 decreased. Second, as the size of PEn decreased, R^2 became smaller. These results are similar to those reported by other research groups (Habier et al., 2007; Habier et al., 2010, Saatchi et al., 2010; Saatchi et al., 2011), who concluded that the closer the relationship between individuals in PEn and those in PEv, the higher the R^2 of GVs. Similarly, using both simulated and real sheep data, Genomic Best Linear Unbiased Prediction was compared with two pedigree based methods (Clark et al., 2012). It was found that both empirical and estimated accuracy of GVs were different for several degrees of relationship. These authors concluded that R^2 of GVs is proportional to the genetic relationship of animals under selection to the reference population. The increase in R^2 of GVs when PEn and PEv are closely related, can be explained by more precise genomic relationships, improving in this way the connectedness between these populations and more distant populations. Accordingly, another research group, Saatchi, Ward and Garrick (2013) concluded that accuracy of GVs deteriorated as the relationship between animals in the PEn and those under selection decreased. One implication of this is that PEn has to be regularly updated to keep the marker effect estimates in sync with new generations of the breeding population (Knol et al., 2016).

On the other hand, as expected, the trend for R^2 held true for PEV, but in the opposite direction. The greater the population size and the closer relationship

between PEn and PEv, the lower PEV. Pszczola, Strabel, Mulder and Calus (2012a) mentioned that PEV can be calculated as the connectedness between the reference population and the animals under evaluation. This may explain the increase in PEV as PEn and PEv became farther apart. Greater connectedness reduces bias, and thus genetic evaluation improves (Kennedy, 1981). The observed trend for PEV held true for both EBV and DEBV response variables.

Table 2 shows the means and their corresponding standard deviations for R^2 and PEV for the combinations of PEn and PEv when the response variable was EBV. In general, R^2 values were only slightly lower than those observed when DEBVs were used as response variables. The trends observed for the decrease in R^2 and the increase in PEV when DEBVs were response variables, as size of PEn diminished and the distance between PEn and PEv augmented, held true for EBVs used as response variables. These results are similar to those obtained by other authors (Calus & Veerkamp, 2007; Meuwissen et al., 2001; Muir, 2007), who found that size of PEn affected R^2 of GVs. The results of the present study and those obtained by other groups of researchers agree with what could be theoretically expected (Daetwyler et al., 2010; Goddard, 2009). These authors developed predictive equations for accuracy of predicted genomic values, which depend on size of PEn, effective population size of the breed, heritability of the trait, and length of genome.

The results of a study with two multi-breed beef cattle populations and Angus and Hereford purebred populations (Weber et al., 2012) used to obtain the GVs and corresponding R^2 for six growth and carcass traits showed that accuracies were lower for prediction equations trained in a single breed. These results were attributed to the smaller number of records derived from a single breed in the training populations. The R^2 range was 0.01 ± 0.10 to 0.65 ± 0.07 , although the authors also reported a negative estimate, -0.10 ± 0.15 .

Our results, regardless of whether DEBV or EBV were used as response variables, are similar to those obtained by Saatchi et al. (2013). These authors evaluated different training populations of Hereford cattle. Accuracy estimates

ranged from 0.15 to 0.52, with 0.30 on average when trained on old animals and validated on young animal populations. The results obtained in our study may be explained by the fact that genomic prediction on closely related individuals is based on relationship; genomic relationships are more accurate when the relationships between PEn and PEv populations are close (Meuwissen et al., 2016). On the other hand, prediction on distant individuals requires DL between QTL and markers (Daetwyler et al., 2012).

The R^2 results of our study are lower than those reported by Pszczola, Mulder and Calus (2011). These authors found that the inclusion of animals with predicted genotypes in the reference population did not significantly increase accuracies of GVs for juvenile animals. They attributed the lack of significance to the low accuracy of predicted genotypes and concluded that inclusion of non-genotyped animals is expected to enhance genomic selection accuracy only when the unknown genotypes can be predicted with high accuracy. The results obtained by these authors varied from 0.57 to 0.96, from 0.48 to 0.88, and from 0.33 to 0.72 for heritabilities of 0.30, 0.05, and 0.01, respectively, under different sizes of the reference population, and different numbers of animals with known or predicted genotypes.

The small difference in GV accuracy that we obtained in our study when EBVs or DEBVs were the response variables agree with reports by other researchers. However, these results are opposite to those observed by Ostersen, Christensen, Henroy, Nielsen, Su and Madsen (2011), who found 18 to 39% higher accuracies for feed conversion ratio and daily gain when they used DEBVs instead of EBVs as response variables. The estimation methodologies they used were GBLUP, Bayesian Lasso, and MIXTURE, where the marker effects are assumed to follow a normal distribution, double exponential, and a mixture of two normal distributions, respectively. The three alternatives of analysis yielded similar reliabilities of the GVs for the two traits analyzed.

Contrary to our results, Ricard, Danvy and Legarra (2013) did not find substantial advantages to genomic values obtained using deregressed EBVs as response

variables or the GBLUP and BayesC π alternatives of analysis compared with conventional BLUP predictions. They followed a specific deregression procedure that included not only the individual's own performance, but also the performance of several relatives (not just offspring), in addition to the genotyped sample. This regression procedure was easy to implement from EBVs, reliabilities, and pedigrees. Unfortunately, accuracy of genomic evaluation, measured by cross validation in several validation samples, was not enough to suggest its use in current breeding plans for the jumping horse population studied. However, the authors mention that this conclusion is related only to accuracy, and the potential benefits of a higher selection intensity, reduced generation intervals, and low inbreeding in the long run should be considered when genomic selection in horses is planned. In dairy cattle similar results were reported by Guo, Lund, Zhang and Su (2010). The authors compared two response variables, EBVs and daughter yield deviations (DYD) on simulated dairy data under eight scenarios of heritability, number of daughters per sire, and number of genotyped sires. They found that DYDs yielded slightly lower reliabilities than EBVs. The average differences in GV accuracy of between EBVs and DYDs were 0.009 for $h^2=0.30$, and 0.035 when $h^2=0.05$.

3.4.2 Genetic correlations between true breeding values and genomic values

Table 3 presents the genetic correlation estimates between true genetic values and the GVs $r_{(TBV,GV)}$ obtained using DEBVs and EBVs as response variables. A slight advantage of using DEBVs, range 0.43 to 0.53, instead of EBVs, range 0.41 to 0.51, held constant throughout all training population sizes. Also, genetic correlation estimates decreased as PEn and PEv separated. The $r_{(TBV,GV)}$ estimates of the present study are higher than those observed by Alarcón-Zúñiga, Ramírez-Flores, Ruíz-Flores, Ramírez-Valverde, Saavedra-Jiménez and Zepeda-Batista (2015), range 0.29 to 0.40, using the same dataset but different models for the genomic analysis. Genetic correlation estimates between direct genomic values and phenotypes from k-fold validation in Red Angus, Angus,

Hereford, Simmental and Limousin ranged from 0.32 to 0.85 for birth weight, weaning weight, milk yield, rib eye muscle area, marbling, direct calving ease, and maternal calving ease (Saatchi et al., 2011; Saatchi et al., 2012). Similarly, genetic correlation estimates between true genetic values and GVs for marbling, using data sets with different proportions of available information, ranged from 0.256 to 0.859 (Spangler et al., 2007). Guo et al. (2010) found genetic correlation estimates between GVs and conventional parent average ranging from 0.457 to 0.688 using three statistical models and eight combinations of heritability and number of daughters per sire.

Some limitations of our study are that a distance between training and evaluation populations needs to be more specific, and size and number of generations in the training population also need to be better determined. Moreover, since our study used simulated information, it does not entirely correspond to real production system conditions.

3.5 Conclusions

The advantage of using deregressed predicted breeding values as the response variable, instead of conventional predicted breeding values, was very slight with any combination of training population size and evaluation generation.

Regardless of the response variable used, predicted genetic value or deregressed predicted genetic value, larger training population were associated with higher genomic values accuracy.

Prediction error variance was low and similar with any combination of training population size and evaluation generation, regardless of the response variable used.

The genetic correlation estimate between true genetic values and genomic values obtained using DEBV as the response variable were slightly higher than those between true genetic values and genomic values obtained using EBV as the response variable.

3.6 Acknowledgements

The authors thank CONACYT for the financial support for the Master of Science studies in Animal Production of the first author.

3.7 Cited Literature

Alarcón-Zúñiga, B., F. Ramírez-Flores, A. Ruíz-Flores, R. Ramírez-Valverde, L. A. Saavedra-Jiménez & J. L. Zepeda-Batista. (2015). Comparación de la exactitud de valores genómicos de animales predichos a través del análisis con dos modelos alternativos. *Agrociencia*, 49, 613-622.

Calus, M. P. L. & R. F. Veerkamp. (2007). Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics*, 124, 362-368.

Clark, S. A., J. M. Hickey, H. D. Daetwyler & J. H. J. van der Werf. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genetics Selection Evolution*, 44, 4-12.

Daetwyler, H. D., K. E. Kemper, J. H. J. van der Werf & B. J. Hayes. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal of Animal Science*, 90, 3375-3384.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva & J. A. Woolliams. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185, 1021-1031.

Fernando, R. & D. J. Garrick. (2009). GenSel – user manual for a portfolio of genomic selection related analyses, create 9.1. http://biggs.ansci.iastate.edu/bigsgui/ReleaseNotes/Version9_1ReleaseNotes.1.pdf. Accessed: January, 2013.

- Garrick, D. J., J. F. Taylor & R. Fernando. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41, 55-62.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis & R. Thompson. (2009). ASREML User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. <http://www.vsnl.co.uk>. Consulted 31st January 2014.
- Goddard, M. E. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, 136, 245-257.
- Guo, G., M. S. Lund, Y. Zhang & G. Su. (2010). Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *Journal of Animal Breeding and Genetics*, 127, 423-432.
- Habier, D., J. Tetens, F. R. Seefried, P. Lichtner & G. Thaller. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, 42, 5-16.
- Habier, D., R. L. Fernando & J. C. M. Dekkers. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177, 2389-2397.
- Hassani, S., M. Saatchi, R. L. Fernando & D. J. Garrick. (2015). Accuracy of prediction of simulated polygenic phenotypes and their underlying quantitative trait loci genotypes using real or imputed whole-genome markers in cattle. *Genetics Selection Evolution*, 47, 99-109.
- Ibanez-Escriche, N. & H. Simianer. (2015). Animal breeding in the genomics era. *Animal Frontiers*, 5, 4-5.
- Kennedy, B. W. (1981). Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. *Journal of Dairy Science*, 64, 689-697.

- Knol, E. F., B. Nielsen & P. W. Knap. (2016). Genomic selection in commercial pig breeding. *Animal Frontiers*, 6, 15-22.
- Meuwissen, T. H. E., B. J. Hayes & M. E. Goddard. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.
- Meuwissen, T., B. Hayes & M. Goddard. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6, 6-14.
- 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, 342-355.
- Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su & P. Madsen. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution*, 43, 38-43.
- Pszczola, M., H. A. Mulder & M. P. L. Calus. (2011). Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. *Journal of Animal Science*, 94, 431-441.
- Pszczola, M., T. Strabel, H. A. Mulder & M. P. L. Calus. (2012a). Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science*, 95, 389-400.
- Pszczola, M., T. Strabel, J. A. M. van Arendonk & M. P. L. Calus. (2012b). The impact of genotyping different groups of animals on accuracy when moving from traditional to genomic selection. *Journal of Dairy Science*, 95, 5412-5421.
- Ricard, A., S. Danvy & A. Legarra. (2013). Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. *Journal of Animal Science*, 91, 1076-1085.

- Saatchi, M., J. Ward & D. J. Garrick. (2013). Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *Journal of Dairy Science*, 91, 1538-1551.
- Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decker, ... & Taylor JF. (2011). Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution*, 43, 40-55.
- Saatchi, M., R. D. Schnabel, M. M. Rolf, J. F. Taylor & D. J. Garrick. (2012). Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genetics Selection Evolution*, 44, 38-49.
- Saatchi, M., S. R. Miraei-Ashtiani, A. Nejati Javaremi, M. Moradi-Shahrehabak & H. Mehrabani-Yeghaneh. (2010). The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *African Journal of Biotechnology*, 9, 438-442.
- Sargolzaei, M. & F. S. Schenkel. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25, 680-681.
- Spangler, M. L., J. K. Bertrand & R. Rekaya. (2007). Combining genetic test information and correlated phenotypic records for breeding value estimation. *Journal of Animal Science*, 85, 641-649.
- Weber, K. L., R. M. Thallman, J. K. Keele, W. M. Snelling, Bennet GL, T. P. L. Smith, ... & L. A. Kuehn. (2012). Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. *Journal of Animal Science*, 94, 4177-4190.
- Weller, J. I., M. Shlezinger & M. Ron. (2005). Correcting for bias in estimation of quantitative trait loci effects. *Genetics Selection Evolution*, 37, 501-522.

Table 1. Mean±standard deviation for accuracy (R^2) and prediction error variance (PEV) of genomic values obtained using deregressed predicted genetic values as response variables, four training populations, and four generations of evaluation.

Evaluation generation	Training population			
	10	9 and 10	8 to 10	7 to 10
R^2				
11	0.52±0.04	0.67±0.03	0.73±0.03	0.77±0.01
12	0.39±0.04	0.55±0.03	0.63±0.03	0.68±0.03
13	0.32±0.05	0.49±0.04	0.58±0.03	0.64±0.03
14	0.28±0.06	0.45±0.04	0.54±0.04	0.60±0.03
PEV				
11	0.05±0.003	0.05±0.003	0.04±0.003	0.04±0.003
12	0.06±0.004	0.06±0.004	0.06±0.004	0.05±0.004
13	0.07±0.005	0.07±0.005	0.06±0.005	0.06±0.005
14	0.07±0.006	0.08±0.006	0.07±0.006	0.07±0.005

Table 2. Mean±standard deviation for accuracy (R^2) and prediction error variance (VEP) of genomic values obtained using predicted genetic values as response variables, four training populations, and four generations of evaluation.

Evaluation generation	Training population			
	10	9 and 10	8 to 10	7 to 10
R^2				
11	0.48±0.04	0.65±0.03	0.71±0.02	0.76±0.02
12	0.38±0.05	0.55±0.03	0.62±0.03	0.68±0.02
13	0.32±0.05	0.49±0.03	0.58±0.03	0.64±0.03
14	0.28±0.06	0.45±0.04	0.54±0.04	0.60±0.03
PEV				
11	0.05±0.003	0.05±0.003	0.04±0.003	0.04±0.003
12	0.05±0.004	0.06±0.004	0.06±0.004	0.05±0.004
13	0.06±0.004	0.07±0.005	0.06±0.004	0.06±0.004
14	0.06±0.005	0.07±0.006	0.07±0.005	0.06±0.005

Table 3. Genetic correlation estimates between true genetic values and genomic values obtained using deregressed predicted genetic values (DEBV) or predicted genetic values (EBV) as response variables, with four training populations.

Training population	Response variable	
	DEBV	EBV
10	0.53	0.51
9 y 10	0.51	0.50
8 a 10	0.48	0.47
7 a 10	0.43	0.41