



UNIVERSIDAD AUTÓNOMA CHAPINGO

**DEPARTAMENTO DE ENSEÑANZA, INVESTIGACIÓN Y
SERVICIO EN ZOOTECNIA**

POSGRADO EN PRODUCCIÓN ANIMAL

EVALUACIÓN GENÉTICA DE BOVINOS SUIZO EUROPEO BASADA EN MARCADORES MOLECULARES

TESIS

**Que como requisito parcial
para obtener el grado de:**

MAESTRO EN CIENCIAS EN INNOVACIÓN GANADERA

Presenta:

SUSANA MORALES PABLO

Bajo la supervisión de: Agustín Ruíz Flores, Ph. D.

y Sergio Pérez Elizalde, Dr.



DIRECCIÓN GENERAL ACADÉMICA
DPTO. DE SERVICIOS ESCOLARES
OFICINA DE EXÁMENES PROFESIONALES



Chapingo, Estado de México, abril de 2018

EVALUACIÓN GENÉTICA DE BOVINOS SUIZO EUROPEO BASADA EN MARCADORES MOLECULARES

Tesis realizada por Susana Morales Pablo bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

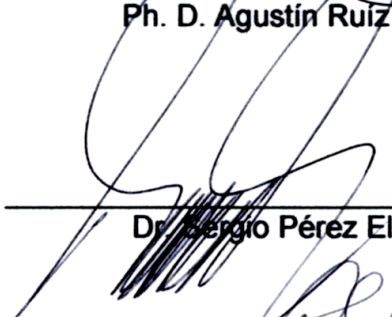
MAESTRO EN CIENCIAS EN INNOVACIÓN GANADERA

Director



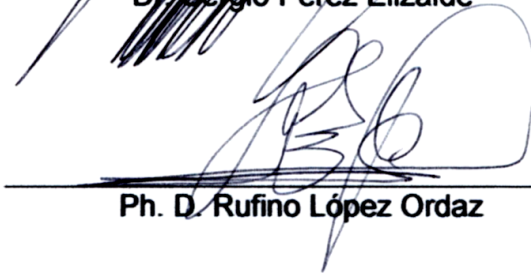
Ph. D. Agustín Ruíz Flores

Co-director



Dr. Sergio Pérez Elizalde

Asesor



Ph. D. Rufino López Ordaz

CONTENIDO

CONTENIDO	III
LISTA DE CUADROS	V
LISTA DE FIGURAS	VII
LISTA DE APÉNDICES.....	VIII
DEDICATORIAS.....	IX
AGRADECIMIENTOS	X
DATOS BIOGRÁFICOS	XI
1. INTRODUCCIÓN GENERAL	1
2. REVISIÓN DE LITERATURA	4
2.1 Predicción cuantitativa basadas en fenotipos y pedigrí	4
2.1.1 Modelo Animal Simple.....	8
2.1.2 Modelo de semental	9
2.1.3 Modelo de semental-vaca	10
2.1.4 Modelo semental-abuelo materno	11
2.1.6 Modelo con efectos ambientes aleatorios	12
2.1.7 Modelo con efectos maternos	13
2.1.8 Modelos para características múltiples	14
2.2 Predicción cuantitativa asistida por marcadores	16
2.3 Tipos de marcadores	17
2.4 Metodologías estadísticas	22
2.4.6 Métodos de regresión penalizada	25
2.4.6.1 Regresión Ridge	26
2.4.6.2 Modelo GBLUP	27
2.4.6.3 Reproducing Kernel Hilbert Space.....	28
2.4.7 Alfabeto Bayesiano	28
2.4.8 Validación cruzada	30

2.5	Herramientas basadas en fenotipos y pedigrí versus asistidas por marcadores en la predicción de valores genéticos	33
2.6	Progreso genético utilizando selección genómica	36
2.7	Perspectivas de selección genómica	39
2.8	Literatura Citada	42
3	PREDICCIÓN DE VALORES GENÉTICOS CON UN MODELO ANIMAL Y BASADOS EN MARCADORES SNP PARA CARACTERÍSTICAS DE CRECIMIENTO EN BOVINOS SUIZO EUROPEO.....	50
3.1	Resumen	50
3.2	Abstract.....	51
3.3	Introducción	52
3.4	Materiales y métodos.....	53
3.4.1	Origen y edición de la información	53
3.4.1.1	Fenotípica.....	53
3.4.1.2	Genómica	54
3.4.2	Análisis de la información.....	55
3.4.2.1	Metodología cuantitativa tradicional (Modelo Animal).....	58
3.4.2.2	Metodologías basadas en marcadores SNP.....	59
3.4.3	Criterios de comparación entre metodologías	60
3.5	Resultados y discusión	61
3.5.1	Magnitud de los estimadores de los parámetros genéticos.....	61
3.5.2	Habilidad de predicción	63
3.5.3	Jerarquización de los VGs predichos de los animales	66
3.5.4	Coeficiente de regresión lineal simple	69
3.6	Conclusiones	72
3.7	Literatura citada	73
3.8	Apéndices	76

LISTA DE CUADROS

Cuadro 1. Marcadores moleculares utilizados en ingeniería genética.	19
Cuadro 2. Genes y condiciones genéticas con efecto conocido en características relacionadas con rendimiento y salud en bovinos.....	21
Cuadro 3. Estadísticos descriptivos, número de observaciones (n) y número de grupos contemporáneos (GC) utilizados en el pre-ajuste de los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo europeo.	54
Cuadro 4. Número y procedencia de los animales utilizados para la obtención de muestras de ADN.....	55
Cuadro 5. Efectos considerados en los modelos de pre-ajuste.	56
Cuadro 6. Estadísticos descriptivos y número de observaciones (n) utilizados en los análisis genéticos para los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo europeo.	58
Cuadro 7. Estimadores de parámetros genéticos para características de crecimiento en bovinos Suizo europeo.	62
Cuadro 8. Promedios de los estimadores de correlación Pearson y cuadrados medios del error (CME) obtenidos para diferentes modelos utilizando validación cruzada de 50 repeticiones para PN, PD y PA de Bovinos Suizo europeo.....	63
Cuadro 9. Estimadores de los coeficientes de correlación Pearson (arriba de la diagonal) y Spearman (debajo de la diagonal), entre paréntesis nivel de significancia para la hipótesis del estimador igual a cero, entre valores genéticos predichos con cuatro modelos de evaluación genética en bovinos Suizo europeo.	67
Cuadro 10. Porcentajes de coincidencia de los 20 animales superiores ordenados respecto a los valores genéticos predichos con cuatro modelos de evaluación genética para pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo europeo.....	68

Cuadro 11. Estimadores de los coeficientes de regresión ($\pm$ error estándar), de los valores fenotípicos pre-ajustados sobre valores genéticos predichos para pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo europeo...	70
Cuadro 12. Estimadores de los coeficientes de regresión ($\pm$ error estándar), utilizando como variable dependiente los valores genéticos predichos mediante un modelo y como variable independiente los valores predichos con los modelos restantes, entre paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, para pesos al nacimiento, al destete y al año en bovinos Suizo europeo.....	71

LISTA DE FIGURAS

Figura 1. Proceso para realizar evaluaciones genéticas y consideraciones para mejorar la predicción de los valores genéticos.	5
Figura 2. Número de registros fenotípicos necesarios para lograr una exactitud determinada del valor genético predicho, dada la heredabilidad de la característica. Se supone un tamaño efectivo de la población (N_e) = 1,000 y una distribución normal de los efectos de QTLs. Fuente: Hayes et al. (2009).	32
Figura 3. Intervalo entre generaciones en bovinos lecheros en un programa de prueba de progenie tradicional.....	33
Figura 4. Cronograma de un programa de cría que utiliza inseminación artificial basado en evaluación genómica.....	34
Figura 5. Número de animales genotipados e incluidos en las evaluaciones genómicas para bovinos lecheros en los EE. UU. desde enero de 2009.	37
Figura 6. Distribución de datos productivos de peso al nacimiento, al destete y al año después de pre-ajustarlos por efectos no genéticos.	57

LISTA DE APÉNDICES

Apéndice A. Rutina en R-3.4.3 para pre-ajuste de registros fenotípicos.....	76
Apéndice B. Rutinas en R-3.4.3 para construcción de matriz X, codificación, imputación y matriz G	79
Apéndice C. Rutina en R-3.4.3 para construcción de matriz A	82
Apéndice D. Rutina en R-3.4.3 para análisis de datos con modelo GBLUP o BLUP con marcadores	83
Apéndice E. Rutina en R-3.4.3 para análisis de datos con modelo GBLUP +A o BLUP con información de marcadores y pedigrí.....	85
Apéndice F. Rutina en R-3.4.3 para análisis de datos con modelo RKHS.....	86
Apéndice G. Validación cruzada con 50 particiones aleatorias	87

DEDICATORIAS

Con amor y cariño para mi familia y amigos

AGRADECIMIENTOS

A la Universidad Autónoma Chapingo y al Posgrado en Producción Animal, por una oportunidad más para fortalecer mi desarrollo profesional.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el financiamiento otorgado para mis estudios de Maestría.

A la Asociación Mexicana de Criadores de Ganado Suizo de Registro, por la información proporcionada para la investigación.

Al Ph. D. Agustín Ruíz Flores, por su excelente dirección en ésta investigación.

A los Ph. D. Sergio Pérez Elizalde y Paulino Pérez Rodríguez por el apoyo y tiempo invertidos para la supervisión y realización de este trabajo.

A los Ph. D. Rufino López Ordaz y José Guadalupe García Muñiz por sus valiosas sugerencias en la revisión de este escrito.

A los M.C. Jonathan, Jorge y Neón por su amistad y apoyo para los análisis.

A cada uno de los que fueron pieza clave para desarrollar y culminar este proyecto.

DATOS BIOGRÁFICOS



Datos personales

Nombre:	Susana Morales Pablo
Fecha de nacimiento	24 de mayo de 1993
Lugar de Nacimiento	Cuetzalan del Progreso, Puebla
CURP	MOPS930524MPLRBS07
Profesión	Ingeniero Agrónomo Especialista en Zootecnia
Cedula Profesional	9926899

Desarrollo Académico

Maestría en Ciencias 2016-2017	Posgrado en Producción Animal Departamento de Enseñanza, Investigación y Servicio en Zootecnia Universidad Autónoma Chapingo
Licenciatura 2011-2015	Departamento de Enseñanza, Investigación y Servicio en Zootecnia Universidad Autónoma Chapingo
Preparatoria 2008-2011	Universidad Autónoma Chapingo

1. INTRODUCCIÓN GENERAL

Los bovinos Suizo europeo son económicamente importantes en México, por su uso como raza pura o en sistemas de cruzamiento para producción de carne, y por su distribución geográfica, en regiones templadas o tropicales (Saavedra-Jiménez et al., 2013). El mejoramiento genético de ésta raza requiere definir criterios de selección que garanticen que los machos y hembras seleccionadas como reproductores posean genes sobresalientes para las características de interés.

En México, se realizan evaluaciones genéticas de bovinos Suizo europeo desde 2003 (Núñez-Domínguez, Ramírez-Valverde, Ruíz-Flores & Domínguez-Viveros, 2003). El objetivo de dichas evaluaciones es predecir el valor genético (VG) de los animales considerando información fenotípica y genealógica propia y de sus parientes. El VG de un individuo como progenitor, se define en términos del comportamiento promedio de su progenie (Henderson, 1984). La adición de registros de un animal o de sus parientes, cambia la predicción de su valor genético predicho, generalmente, mediante el incremento de la confiabilidad de la predicción (Ramírez-Valverde, Núñez-Domínguez, Ruíz-Flores, García-Muñiz & Domínguez-Viveros, 2008a). La mitad del VG de un animal para una característica de interés es lo que el animal transmite a su progenie.

Con los avances en bioinformática y genética molecular, la selección genómica (SG), como una nueva técnica de selección genética, se desarrolló a partir de 2001 (Meuwissen, Hayes & Goddard, 2001). La SG combinada con los programas de mejoramiento existentes, proporciona un nuevo nivel de información que permite identificar y seleccionar animales con mayor exactitud a edades tempranas de los animales candidatos a selección.

Con el uso de información de los genotipos en términos de SNPs (Nucleótidos de Polimorfismo Simple), es posible predecir el VG de los animales (Meuwissen et al., 2001), el rastreo de genes relacionados con características de interés y trazabilidad y pruebas de paternidad en una población a edad temprana. Esto ha

permitido que la predicción genómica sea una tecnología viable en el mejoramiento genético de plantas y animales (de los Campos, Hickey, Pong-Wong, Daetwyler & Calus, 2013; Hayes, Bowman, Chamberlain & Goddard, 2009; VanRaden et al., 2009), al disminuir los costos de los programas de selección de animales y el intervalo entre generaciones, y aumentar la exactitud de los valores genéticos predichos, además, de complementar la evaluación genética tradicional basada en pedigrí y fenotipos.

La predicción de los VG puede realizarse con una exactitud relativamente alta, por lo que se requiere que un número grande de SNPs cubran el genoma completo. La predicción por modelos convencionales de evaluación genética, considerando toda la información de marcadores, no es posible, es necesario un enfoque diferente para usar de manera eficiente toda la información genómica. Esto puede realizarse mediante alternativas como GBLUP, Regresión Ridge-BLUP (RR-BLUP), RKHS, Lasso Bayesiano, BayesA, BayesB y BayesC, entre otros (de los Campos et al., 2013). Éstas metodologías pueden diferir en magnitud y exactitud en la predicción de los valores genéticos.

El alcance que tienen las metodologías en la predicción de los VG, depende en gran medida de la calidad y estructura de los datos fenotípicos y de factores como: pre-ajuste de la información por efectos no genéticos, control de calidad de SNPs, tamaño de la población de referencia, relaciones genéticas entre población de entrenamiento y validación y estructura genética de las características (Habier et al., 2007; Daetwyler, Villanueva & Woolliams, 2008; Meuwissen et al., 2001; VanRaden & Sullivan 2010).

Los estudios con información real han confirmado la superioridad de SG con respecto a las predicciones basadas sólo en fenotipos y pedigrí. El caso extraordinario ocurre en el ganado lechero Holstein Friesian en donde varios estudios (Hayes et al., 2009; VanRaden et al., 2009) han confirmado que SG puede alcanzar una predicción más alta que la basada sólo en el pedigrí. El potencial de SG también se ha confirmado en bovinos productores de carne

(Garrick, 2011; Saatchi et al., 2011), ovinos (Daetwyler et al., 2010), pollos (González-Recio et al., 2008) y gallinas (Wolc et al., 2011a,b).

Por lo anterior, el objetivo del trabajo de investigación fue comparar las predicciones de valores genéticos para características de crecimiento obtenidas con el modelo animal y con modelos basados en marcadores SNP en bovinos Suizo europeo.

En el primer capítulo de ésta tesis, se revisó de manera breve la importancia de las metodologías cuantitativas y la SG para predecir los VG y los antecedentes de las mismas a nivel nacional.

En el segundo capítulo de ésta tesis se revisó literatura científica relacionada a la evaluación genética tradicional y asistida por marcadores moleculares. En él se describen y comparan metodologías cuantitativas con información fenotípica y de pedigrí y asistidas por marcadores moleculares para predecir los VG, algunos beneficios de la SG y sus perspectivas.

En el tercer capítulo, se presenta un trabajo de investigación donde se comparó una metodología cuantitativa tradicional con algunas basadas en marcadores SNPs en la predicción de los valores genéticos de pesos al nacimiento, al destete y al año en bovinos Suizo europeo.

2. REVISIÓN DE LITERATURA

Los animales candidatos a reproductores se pueden seleccionar por su valor genético (VG) predicho. El VG es el valor de un individuo como progenitor y se define en términos de comportamiento promedio de su progenie (Henderson, 1984). La mitad del VG de un animal para una característica de interés es lo que un progenitor transmite a su progenie. La predicción exacta del VG es crucial para caracterizar el ganado de acuerdo con las influencias genéticas y no genéticas, las cuales deben considerarse en el modelo de evaluación.

2.1 Predicción cuantitativa basadas en fenotipos y pedigrí

El mejoramiento genético requiere criterios de selección donde los individuos seleccionados como futuros reproductores; posean genes que influyan positivamente sobre las características de interés. Uno de estos criterios es la obtención de una predicción del VG de los animales a través los predictores lineales insesgados mejores (BLUP). Es importante considerar información de sus registros y sus parientes, mediante metodologías de modelos lineales. Las predicciones de los VG, se reportan como Diferencias Esperadas en la Progenie (DEP) y representan la mitad del VG del animal, en otras palabras, es la proporción de genes que los padres transmiten a su progenie (Núñez-Domínguez, Ramírez-Valverde, Ruíz-Flores & Moreno-Hidalgo, 2014). Las DEP sirven para predecir el comportamiento futuro de la progenie de un animal comparado con la de otros animales dentro de la población bajo evaluación para una característica y año específico. La Figura 1 muestra cómo se realizan las evaluaciones genéticas utilizando herramientas cuantitativas.

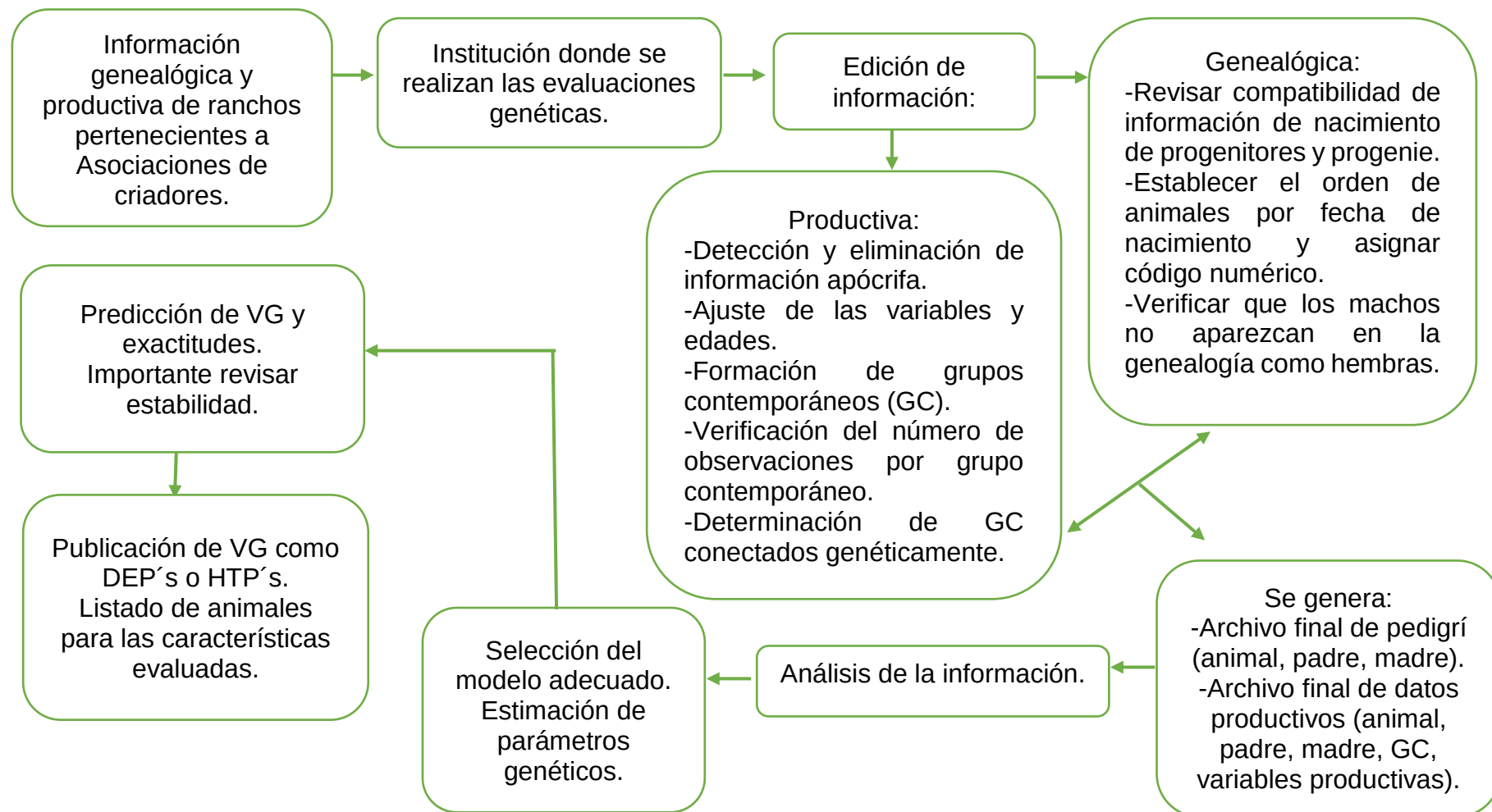
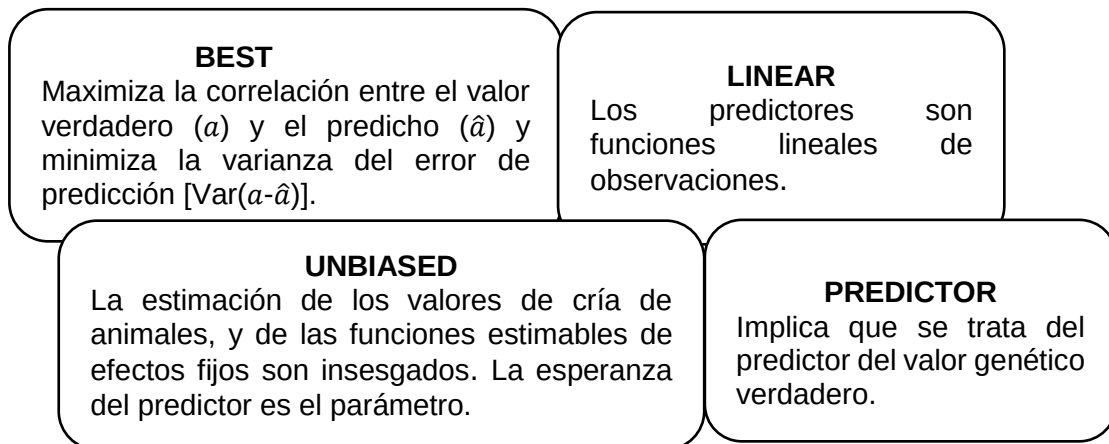


Figura 1. Proceso para realizar evaluaciones genéticas y consideraciones para mejorar la predicción de los valores genéticos.

Los modelos mixtos permiten obtener el predictor lineal insesgado mejor (BLUP, Best Linear Unbiased Predictor) del VG de los animales evaluados en función de sus registros y parientes. Estos predictores minimizan los cuadrados medios del error. En otras palabras, minimizan la varianza del error de predicción (VEP). Las ecuaciones de modelos mixtos permiten el cómputo simultáneo de los mejores estimadores lineales insesgados mejores (BLUE) de los efectos fijos y de los predictores lineales insesgados mejores (BLUP) de valores genéticos (Martínez-Niño, Manrique-Perdomo & Elzo, 2012).

Las propiedades de BLUP están incorporadas en su nombre:



Henderson (1984) mencionó que la metodología de modelos mixtos para BLUP constituye una herramienta potente y flexible para predicciones con propiedades muy deseables:

1. La varianza del error de predicción es pequeña comparada con la de cualquier otro predictor; por ejemplo, cuadrados mínimos o índice de selección.
2. La correlación entre el predictor y el valor verdadero es mayor que la obtenida con otros métodos.
3. La probabilidad de jerarquización correcta en los valores genéticos de los animales, se maximiza.
4. Las alteraciones en la variabilidad genética de las poblaciones (debidas a consanguinidad y selección) se toman en cuenta.

5. Permite predecir los VG de animales sin registros fenotípicos.
6. Los VG se predicen considerando efectos fijos y viceversa (por lo que pueden compararse animales de diferentes hatos), y son relativos a una generación o año base.
7. Cuando se utiliza el modelo animal para la predicción de los VG, éstos se ajustan por el mérito genético de la pareja con la que se aparean los animales.

Sin embargo, la metodología BLUP no corrige la calidad y confiabilidad de los datos, tampoco el tratamiento preferencial de ciertos animales, ni heterogeneidad de varianzas. Posteriormente, de acuerdo con Interbull (2017), esto fue superado con la metodología MACE (Multi-trait Across Country Evaluation), desarrollada por Schaeffer (1994).

Los grupos contemporáneos (GC) se utilizan en las evaluaciones genéticas para reducir sesgos en la predicción de los VG debido a diferencias en el manejo y condiciones ambientales de crianza de los animales. En evaluaciones genéticas de bovinos productores de carne, se considera al GC como efecto fijo. Ramírez-Valverde, Núñez-Domínguez, Ruíz-Flores, García-Muñiz y Magaña-Valencia (2008b) indicaron que en las evaluaciones genéticas nacionales de bovinos Suizo europeo en México, los grupos contemporáneos deben formarse con un mínimo de siete registros e incluirse en el modelo de análisis como efectos fijos, esto resulta en mayores estimadores de heredabilidad, con diferencias marginales en exactitud y jerarquización de valores genéticos.

Un modelo animal es un modelo mixto donde el efecto aleatorio es el animal con registros, y donde se pueden incluir efectos aleatorios adicionales como: genéticos maternos, de ambiente materno común o miembros de una familia criados con condiciones similares, de ambiente permanente, de semental y de interacciones, por ejemplo: semental x hato.

Los modelos animales pueden ser diferentes según la situación biológica y estructura de los datos; un criterio importante en la selección de modelos, es su habilidad de predicción que puede verificarse mediante métodos como la

validación cruzada (VC). La VC consiste en omitir aleatoriamente una parte de datos, ajustar el modelo a los remanentes y probar su ajuste en la parte de información omitida (Shao, 1993; citado por Ramírez et al., 2007). Los estimadores de los coeficientes de correlación proporcionan información comparativa del comportamiento de los modelos, lo que es útil para su jerarquización (Ramírez-Valverde, Hernández-Álvarez, Núñez-Domínguez, Ruíz-Flores & García-Muñiz, 2007). Una mayor correlación entre muestras complementarias implica una mayor estabilidad del modelo para obtener soluciones de los VG de los animales con registros eliminados (Ramírez-Valverde, Misztal & Bertrand, 2001). Por tanto, se pueden hacer suposiciones con relación a la estructura de los datos y utilizar alguno de los modelos que se describen en los siguientes apartados.

2.1.1 Modelo Animal Simple

El modelo animal simple considera los efectos genéticos directos de los animales. La ventaja principal que ofrece es que los animales sin registros pueden evaluarse a través de las relaciones genéticas de parentesco. La información disponible se puede usar para mejorar la exactitud de las predicciones.

El modelo en términos de matrices es:

$$y = X\beta + Zu + \varepsilon \quad [1.1]$$

donde:

y : es el vector $nx1$ de observaciones fenotípicas, n : número de observaciones,

β : es el vector $px1$ de efectos fijos, p : niveles de los efectos fijos en el modelo,

u : es el vector $qx1$ de efectos aleatorios del animal, q : número de niveles para el efecto aleatorio; i. e. número de animales,

ε : vector $nx1$ de efectos aleatorios residuales,

X : matriz de incidencia $n \times p$, que relaciona los registros fenotípicos con los efectos fijos correspondientes, y

Z : matriz de diseño $n \times q$, que relaciona los registros fenotípicos con los correspondientes efectos aleatorios del animal.

En el modelo [1.1] se supone que.:

$$E(y) = X\beta; \text{ y } E(u) = E(e) = 0$$

La suposición es, que los efectos residuales, que incluyen otros efectos aleatorios ambientales y no aditivos, se distribuyen idéntica e independientemente con media cero y varianza σ_e^2 , por tanto:

$$\text{var}(e) = I\sigma_e^2 = R, \text{ var}(a) = A\sigma_a^2 = G \text{ y } \text{cov}(a,e) = \text{cov}(e,a) = 0.$$

Henderson (1950) indicó las ecuaciones de modelos mixtos (EMM) para estimar los efectos fijos (β) y predecir los efectos aleatorios (u) simultáneamente.

Las EMM para [1.1] son:

$$\begin{bmatrix} X^t R^{-1} X & X^t R^{-1} Z \\ Z^t R^{-1} X & Z^t R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^t R^{-1} y \\ Z^t R^{-1} y \end{bmatrix}$$

Suponiendo que R y G son no-singulares, $G^{-1} = A^{-1}(1/\sigma_a^2)$, $R^{-1} = I(1/\sigma_e^2)$ y **A** es la matriz de relaciones genéticas aditivas entre animales en el pedigrí.

Puesto que R^{-1} es una matriz identidad, previamente definida en ésta sección, se puede factorizar en ambos lados de la ecuación, y si, posteriormente se multiplica por σ_e^2 , el resultado es el sistema de ecuaciones más familiar:

$$\begin{bmatrix} X^t X & X^t Z \\ Z^t X & Z^t Z + A^{-1} \lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^t y \\ Z^t y \end{bmatrix}$$

donde:

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2}{h^2}$$

Las soluciones $\hat{\beta}$ y $\hat{u}$ serán BLUE y BLUP, respectivamente.

2.1.2 Modelo de semental

El modelo de evaluación es sólo de sementales, usando información de su progenie.

El modelo en términos de matrices es:

$$y = X\beta + Zu + e \tag{1.2}$$

donde:

los términos similares a los del modelo animal, se definen de igual manera, y ahora:

Z : matriz de diseño nxq , que relaciona los registros fenotípicos con los sementales,

u : vector $qx1$ de efectos aleatorios de sementales, q es el número de sementales,

Las EMM para [1.2] son:

$$\begin{bmatrix} X^t X & X^t Z \\ Z^t X & Z^t Z + A^{-1} \lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^t y \\ Z^t y \end{bmatrix}$$

A es ahora la matriz de parentescos entre sementales y

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2}$$

La ventaja principal del modelo de semental comparado con el modelo animal es la reducción en el número de ecuaciones, sólo los sementales se evalúan (Mrode & Thompson, 2005). Sin embargo, esto constituye su principal desventaja, sólo los sementales se evalúan y se desaprovechan las relaciones de parentesco entre todos los animales en el pedigrí. Éste modelo asume que los sementales se aparean con hembras de similar mérito genético, esto es porque el valor genético de ellas no es considerado, si existe apareamiento preferencial puede provocar sesgo en los VG. Adicionalmente, no considera la estimación de efectos maternos.

2.1.3 Modelo de semental-vaca

El modelo considera los efectos genéticos directos de sementales y vacas.

El modelo escrito con matrices es:

$$y = X\beta + Z_1 s + Z_2 d + e \quad [1.3]$$

donde:

Z_1 : matriz de diseño nxp , la cual relaciona los registros con los sementales,

s : es el vector $px1$ de efectos aleatorios de sementales,

Z_2 : matriz de diseño nxp , que relaciona los registros con las vacas, y

d : es el vector $px1$ de efectos aleatorios de vacas.

Los demás términos en el modelo se definen como se describe previamente.

Las EMM para [1.3] son:

A es ahora la matriz de parentescos para sementales.

$$\begin{bmatrix} X^t X & X^t Z_1 & X^t Z_2 \\ Z_1^t X & Z_1^t Z_1 + A^{ss} \lambda^{ss} & Z_1^t Z_2 + A^{sd} \lambda^{sd} \\ Z_2^t X & Z_2^t Z_1 + A^{ds} \lambda^{ds} & Z_2^t Z_2 + A^{dd} \lambda^{dd} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{s} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} X^t y \\ Z_1^t y \\ Z_2^t y \end{bmatrix}$$

$$\lambda^{ss} = \frac{\sigma_e^2}{\sigma_s^2}, \sigma_s^2 = \frac{1}{4} \sigma_a^2$$

$$\lambda^{dd} = \frac{\sigma_e^2}{\sigma_d^2}, \sigma_d^2 = \frac{1}{4} \sigma_a^2 + \sigma_m^2 + \sigma_{am}$$

$$\lambda^{sd} = \frac{\sigma_e^2}{\sigma_{sd}}$$

2.1.4 Modelo semental-abuelo materno

De acuerdo con Elzo (2014) se asume que:

- Los padres no tienen registros propios, y
- Las madres están relacionadas sólo través de sus padres.

El modelo en términos matriciales es:

$$y = X\beta + Z_1 s + Z_2 mgs + e \quad [1.4]$$

donde:

Z_1 : matriz de diseño $n \times p$, que relaciona los registros con los sementales,

s : es el vector $p \times 1$ de efectos aleatorios de sementales,

Z_2 : matriz de diseño $n \times p$, que relaciona los registros con los abuelos maternos,

mgs : es el vector $p \times 1$ de efectos aleatorios de los abuelos maternos.

Las EMM para [1.4] son:

$$\begin{bmatrix} X^t X & X^t Z_1 & X^t Z_2 \\ Z_1^t X & Z_1^t Z_1 + A^{ss} \lambda^{ss} & Z_1^t Z_2 + A^{s,mgs} \lambda^{s,mgs} \\ Z_2^t X & Z_2^t Z_1 + A^{mgs,s} \lambda^{mgs,s} & Z_2^t Z_2 + A^{mgs,mgs} \lambda^{mgs,mgs} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{s} \\ \widehat{mgs} \end{bmatrix} = \begin{bmatrix} X^t y \\ Z_1^t y \\ Z_2^t y \end{bmatrix}$$

A es ahora la matriz de parentescos para sementales y

$$\lambda^{ss} = \frac{\sigma_e^2}{\sigma_s^2},$$

$$\lambda^{s,mgs} = \frac{\sigma_e^2}{\sigma_{s,mgs}},$$

$$\lambda^{mgs,mgs} = \frac{\sigma_e^2}{\sigma_{mgs}^2}, \sigma_{mgs}^2 = \frac{1}{16} \sigma_a^2 + \frac{1}{4} \sigma_m^2 + \frac{1}{4} \sigma_{am}^2$$

2.1.6 Modelo con efectos ambientes aleatorios

Elzo (2014) mencionó que los objetivos de este modelo son:

- Predecir los VG de los individuos, $\hat{u}$, basándose en sus propios registros y (o) en los registros de sus familiares.
- Predecir el efecto ambiental común a todos los registros de un animal o entre miembros de una familia criados juntos. El modelo con efectos ambientes aleatorios o de repetibilidad se utiliza principalmente en cerdos y aves porque los individuos comparten un ambiente común y los efectos se asocian con hermanos completos o medios hermanos maternos.
- Predecir el efecto ambiental entre registros de un individuo, debido a factores ambientales que afectan permanentemente los registros del animal. Por ejemplo, en el peso al destete, se puede incluir un efecto permanente ambiental para contabilizar pérdidas en PL causadas por mastitis crónica o un cuarto seco o dañado permanentemente.

La "Habilidad Real de Producción" o "Capacidad Probable de Producción" de un animal se define como la suma del VG y los efectos ambientales permanentes para un animal, esto es, $\hat{v}_i = \hat{u}_i + \widehat{PE}_i$.

La repetibilidad es la correlación entre registros de un individuo.

$$r_e = \frac{\sigma_g^2 + \sigma_{PE}^2}{\sigma_g^2 + \sigma_{PE}^2 + \sigma_{TE}^2}$$

El modelo en términos de matrices es:

$$y = X\beta + Zu + W_{PE} + e \quad [1.5]$$

Las EMM para [1.5] son:

A es ahora la matriz de parentescos para sementales.

$$\begin{bmatrix} X^tX & X^tZ & X^tW \\ Z^tX & Z^tZ + A^{-1}\lambda & Z^tW \\ W^tX & W^tZ & W^tW + I\gamma \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{s} \\ \widehat{PE} \end{bmatrix} = \begin{bmatrix} X^ty \\ Z^ty \\ W^ty \end{bmatrix}$$

$$\text{donde: } \lambda = \frac{\sigma_e^2}{\sigma_a^2} \text{ y } \gamma = \frac{\sigma_e^2}{\sigma_{PE}^2}$$

Los factores ambientales pueden ser parte importante de la covarianza entre individuos, al contabilizar dichos efectos puede mejorarse la exactitud en la predicción de VG.

El modelo de repetibilidad es apropiado cuando se registran medidas múltiples de la misma característica en un individuo.

2.1.7 Modelo con efectos maternos

En bovinos productores de carne, se tiene la particularidad de que algunas características son influidas por efectos maternos. Esto es, existe un efecto debido a la madre del animal y de ésta manera el desempeño del mismo depende de los efectos genéticos directos y del ambiente provisto por la madre (Mrode & Thompson, 2005). El efecto materno al no considerarse en el modelo, sesgará la predicción del VG de los animales. Ejemplos de este tipo de características son las de crecimiento, medidas hasta el destete.

El modelo en términos de matrices es:

$$y = X\beta + Zu + Wm + S_{PE} + e \quad [1.6]$$

donde:

W : matriz de incidencia $n \times p$, que relaciona las observaciones con los efectos genéticos maternos,

m : es el vector de efectos genéticos maternales (aleatorios), y

los demás términos se definen como anteriormente.

La estructura de varianzas y covarianzas para los efectos aleatorios es la siguiente:

$$V \begin{bmatrix} u \\ m \\ PE \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & A\sigma_{am} & 0 & 0 \\ A\sigma_{am} & A\sigma_m^2 & 0 & 0 \\ 0 & 0 & I\sigma_{PE}^2 & 0 \\ 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

donde:

σ_a^2 es la varianza genética aditiva directa, σ_m^2 es la varianza genética aditiva materna,

σ_{am} es la covarianza entre efectos genéticos aditivos directo y materno,

σ_{PE}^2 es la varianza de efectos de ambiente permanente de la madre, y

σ_e^2 es la varianza residual.

Las EMM para [1.6] son:

$$\begin{bmatrix} X^t X & X^t Z & X^t W & X^t S \\ Z^t X & Z^t Z + A^{-1} \lambda_1 & Z^t W + A^{-1} \lambda_2 & Z^t S \\ W^t X & W^t Z + A^{-1} \lambda_3 & W^t W + A^{-1} \lambda_4 & W^t S \\ S^t X & S^t Z & S^t W & S^t S + I_Y \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{m} \\ \widehat{PE} \end{bmatrix} = \begin{bmatrix} X^t y \\ Z^t y \\ W^t y \\ S^t y \end{bmatrix}$$

donde: $\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2}$, $\lambda_2 = \frac{\sigma_e^2}{\sigma_{am}}$, $\lambda_3 = \frac{\sigma_e^2}{\sigma_{am}}$, $\lambda_4 = \frac{\sigma_e^2}{\sigma_m^2}$, y $\gamma = \frac{\sigma_e^2}{\sigma_{PE}^2}$.

2.1.8 Modelos para características múltiples

En la práctica, la selección de los animales genéticamente superiores es una combinación de varias características, de modo que, el modelo para características múltiples es el adecuado para analizar y aprovechar las correlaciones genéticas y fenotípicas entre ellas.

El modelo en términos de matrices es:

$$y = X_i \beta_i + Z_i u_i + e_i \quad (i = 1, \dots, t) \quad [1.7]$$

donde:

i : es el número de características a evaluar,

Para dos características el modelo en forma matricial es:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Las EMM para [1.7] son:

$$\begin{bmatrix} X_1^t R^{11} X_1 & X_1^t R^{12} X_2 & X_1^t R^{11} Z_1 & X_1^t R^{12} Z_2 \\ X_2^t R^{21} X_1 & X_2^t R^{22} X_2 & X_2^t R^{21} Z_1 & X_2^t R^{22} Z_2 \\ Z_1^t R^{11} X_1 & Z_1^t R^{12} X_2 & Z_1^t R^{11} Z_1 + A^{-1} g^{11} & Z_1^t R^{12} Z_2 + A^{-1} g^{12} \\ Z_2^t R^{21} X_1 & Z_2^t R^{22} X_2 & Z_2^t R^{21} Z_1 + A^{-1} g^{21} & Z_2^t R^{22} Z_2 + A^{-1} g^{22} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} X_1^t R^{11} y_1 + X_1^t R^{12} y_2 \\ X_2^t R^{21} y_1 + X_2^t R^{22} y_2 \\ Z_1^t R^{11} y_1 + Z_1^t R^{12} y_2 \\ Z_2^t R^{21} y_1 + Z_2^t R^{22} y_2 \end{bmatrix}$$

donde: $g^{11} = \sigma_{a_1}^2$, $g^{12} = \sigma_{a_1 a_2}$, $g^{21} = \sigma_{a_2 a_1}$, $g^{22} = \sigma_{a_2}^2$

La principal ventaja del análisis multivariado incluye un incremento en la precisión de las evaluaciones. La ganancia en exactitud depende de la diferencia absoluta

entre las correlaciones genéticas y residuales entre las características. Las diferencias mayores en éstas correlaciones, producen un incremento en la precisión de las evaluaciones (Schaeffer, 1984). Por ejemplo, si la heredabilidad y las correlaciones genéticas y ambientales para dos características son iguales, las predicciones multivariadas son equivalentes a las del análisis univariado para cada característica.

Las matrices de incidencia o de diseño pueden ser iguales o diferentes para las características (Mrode, 2005). Las matrices de diseño iguales para todas las características implican que todos los efectos en el modelo afectan todas las características y no existen registros faltantes. Las matrices de incidencia iguales con registros perdidos aplican cuando todas las características en un análisis multivariado no se observan en todos los animales. Por ejemplo, en análisis univariados de características como PD y PA, no se considera que después del destete ya fueron desechados algunos animales. Sin embargo, utilizando análisis bivariados se pueden predecir los VG para PA para todos los animales, no únicamente para los animales con registro de PD. Las matrices de diseño diferentes surgen cuando las características se ven afectadas por diferentes efectos fijos o aleatorios. Un ejemplo en el que los efectos aleatorios pueden ser diferentes es el análisis conjunto para el peso al destete y al año de bovinos productores de carne. El efecto materno sólo es importante para el peso al destete y el modelo para el análisis incluirá los efectos maternos sólo para ésta característica.

Ramírez-Valverde et al. (2007) compararon análisis uni- y multivariados de peso al nacimiento (PN), al destete (PD) y al año (PA) en la evaluación genética de bovinos Angus y Tropicarne, donde los criterios de comparación fueron: la magnitud de parámetros genéticos, habilidad de predicción de los modelos, y exactitud de los VG predichos. Para PD y PA, los análisis bivariados mejoraron los estimadores de heredabilidad directa y materna, la exactitud de las predicciones de los valores genéticos, y la habilidad de predicción de los modelos, no se observaron ventajas al usar análisis trivariados. Además, el uso

de análisis multivariados con respecto a univariados cambió la jerarquización de los animales en casi todas las variables, por lo que, Ramirez-Valverde y colaboradores (2007) recomiendan para Angus un análisis individual para PN, y bivariado para PD y PA; mientras que en Tropicarne, recomiendan un análisis bivariado de PN y PD, y para PD y PA el análisis bivariado de las dos características.

Hasta este momento se ilustraron diferentes modelos que utilizan información fenotípica y de pedigrí para obtener el mérito genético de un animal. La confiabilidad de las predicciones de valores genéticos dependerá de que tan bien se represente cada situación biológica en los modelos y cuan confiable sea la información productiva y genealógica.

2.2 Predicción cuantitativa asistida por marcadores

La genómica estudia la caracterización de los genomas completos, esto es, la localización de los genes en los cromosomas, la construcción de mapas genéticos y la secuenciación completa de los genomas y las proteínas (Solarte-Portilla & Rosero-Galindo, 2008).

Con base en los avances en genética molecular y bioinformática, una nueva técnica de cría se desarrolló a principios de los años 2000: la selección genómica (SG) (Hannachi & Tichit, 2016). La SG se empezó a implementar a partir de la publicación del estudio de Theo Meuwissen, Ben Hayes y Mike Goddard (Meuwissen et al., 2001).

La asociación entre ingeniería genética (secuenciación del genoma) y bioinformática (procesamiento de datos) dieron como resultado nuevos métodos de selección, y en combinación con los programas de mejoramiento genético tradicional, donde la recolección de información genealógica y fenotípica es rutinaria, proporcionó un nivel nuevo de información para identificar y seleccionar los animales genéticamente superiores.

Para implementar la SG es necesario enviar muestras al laboratorio para su genotipado, ahí se extrae el ADN (Ácido Desoxirribonucleico) de la muestra. En 2015, las principales fuentes de ADN enviadas a los laboratorios incluyeron tejido (64%), pelo (31%), sangre (2%), frotados nasales (<1%) y semen (<1%) (Wiggans, Cole, Hubbard & Sonstegard, 2017). La amplificación y fragmentación del ADN, la hibridación con un chip de genotipado, el marcado y la detección del genotipo duran aproximadamente tres días. Los datos generados a partir de un lector láser se agrupan para determinar los genotipos de SNP. El uso de ésta información depende del fin que se le desee dar, como: predicción de valores genéticos, rastreo de genes relacionados con características de interés, trazabilidad, pruebas de paternidad, etc.

2.3 Tipos de marcadores

El conocimiento del genoma de especies domésticas ha permitido la selección de características productivas y relacionadas a la salud de los animales para la aplicación de técnicas moleculares en mejoramiento genético.

El genoma de *Bos taurus* se estima en aproximadamente 2,697.56 Mb (millones de pares de bases), organizado en 29 pares de cromosomas autosómicos y dos sexuales. Los cromosomas bovinos son acrocéntricos, contienen un mínimo de 22,000 genes que codifican proteínas y 41.7% de las bases G y C (NCBI, 2017). Algunas características se controlan por un gen único (características monogénicas o cualitativas), mientras que otros se controlan por muchos genes (características poligénicas o cuantitativas).

Los estudios de asociación, que consideran todo el genoma (GWAS, Genome-Wide Association Studies) y la selección asistida por marcadores (MAS, *Marker Assisted Selection*), son posibles debido a la disponibilidad de tecnologías que permiten el genotipado de alto rendimiento de marcadores moleculares.

Un marcador molecular es un segmento de ADN de longitud variable con ubicación específica en un cromosoma. Existen diferentes tipos de marcadores

(Cuadro 1), uno de los que se ha utilizado más en GWAS y MAS son los llamados polimorfismos de un nucleótido sólo (SNPs). Estos SNPs son variantes, o alelos, en la secuencia de ADN que pueden estar asociados con la expresión de una característica en alguna especie. Los SNPs pueden localizarse en regiones codificantes, reguladoras y no codificantes. Cuando están presentes en regiones codificantes, muestran mayor asociación con algún carácter de interés (Martínez, Helguera & Carrera, 2011).

En el Cuadro 2 se presentan las regiones genómicas asociadas con características productivas, enfermedades infecciosas y genéticas en bovinos productores de carne y leche (Casas & Kehrlí, 2016). Recientemente los proyectos de secuenciación del genoma han descubierto un gran número de SNPs, lo que favoreció que la selección genómica se esté implementado en algunas especies domésticas.

Cuadro 1. Marcadores moleculares utilizados en ingeniería genética.

Tipo de marcador	Principal uso	Requerimientos	Características
Polimorfismo de base única (SNP)	Mapas genéticos	ADN clonado y secuencias de DNA	Este tipo de marcador se diferencia en una simple posición nucleotídica, tales como sustitución, delección o inserción de un sólo nucleótido. Su máxima heterocigosidad es 50%. El nivel de información, facilidad y costo de detección de los marcadores varían dependiendo del tipo de marcador utilizado.
Polimorfismo de longitud de fragmentos de restricción (RFLP)	Mapas genéticos	ADN clonado	El genoma total de un individuo, es cortado por una endonucleasa de restricción, generándose una serie de fragmentos cuya longitud se determina por el reconocimiento de una secuencia de nucleótidos. Los cambios nucleotídicos generados por distintos cortes permiten detectar el polimorfismo entre dos o más individuos. Dicho polimorfismo es evidente mediante geles de electroforesis que permitan separar fragmentos de tamaño diferentes.
Mini satélites o VNTR número de secuencias de tamaño variable	Huellas digitales	Secuencias repetitivas y pruebas de hibridación	Son marcadores polimórficos que tienen altas tasas de heterocigosis en las poblaciones.
Micro satélites o secuencias repetidas cortas (SSR)	Pruebas de paternidad, identificación de individuos, mapas genéticos, genómica comparativa, estudios de genética poblacional, asignación de individuos a una raza	ADN clonado y secuencias de ADN	Detección de fragmentos específicos de ADN y da la medida de los alelos en pares de bases en cada una de las regiones, son fácil de medir y analizar, son muy confiables, repetitivos y automatizables.

Polimorfismo amplificado aleatoriamente (RAPD)	Huellas digitales	Oligonucleótidos cortos aleatorios	Se utilizan cebadores cortos (10 pares de bases) que se utilizan de forma individual en reacciones de PCR con un nivel de especificidad bajo, de ésta forma se genera un patrón de bandas sencillo que dependerá de la capacidad de los cebadores para encontrar zonas en el ADN, parcial o totalmente complementarias, en función de la secuencia nucleotídica, la amplificación evidenciará el polimorfismo existente entre dos individuos, dando lugar o no a determinadas bandas (Avisé, 2004). Tiene repetibilidad baja en los análisis, subestima la cantidad de polimorfismos existentes.
Polimorfismo de longitud de fragmentos amplificados (AFLP)			Estos marcadores se consiguen al combinar dos metodologías: enzimas de restricción en cadena de la polimerasa, lo que genera patrones de bandas complejos en donde es fácil encontrar polimorfismos claramente diferenciables (Avisé, 2004).

Fuente: Dodgson et al. (1997); y Aranguren-Méndez et al., 2000; citados por Cuevas, 2013).

Cuadro 2. Genes y condiciones genéticas con efecto conocido en características relacionadas con rendimiento y salud en bovinos.

Condición	Cromosoma	Locus ^z	Gen ^y
Deficiencia de adhesión de leucocitos bovinos	BTA1	BLAD	CD18
Deficiencia del Uridin monofosfato sintasa	BTA1	DUMPS	UMPS
Displasia tubular renal	BTA1		CLDN16
Nefritis crónica intersticial bovina	BTA1	CINF	PCLN1/CL16
Desarrollo de cuernos	BTA1	POLL	HSA21
Miostatina	BTA2	mh	MSTN
Malformación vertebral compleja	BTA3	CVM	SLC35A3
Leptina	BTA4	LEP	LEP
Osteopetrosis	BTA4	-	SCL4A2
Factor 1 activador del péptido de apoptosis	BTA5	HH1	APAF1
Color de pelo	BTA5		KITLG
Enanismo condrodisplástico	BTA6	BCD	LBN
Aracnomelia	BTA5		SUOX
Caseína	BTA6	CN	CSN
Calpastatina	BTA7	CAST	CAST
Hernia umbilical	BTA8	UH	-
Pérdidas embrionarias	BTA8	-	SMC2
Lactoglobulina	BTA11	-	SMC2
Citrulinemia	BTA11	ASS	ASS
Deficiencia de colesterol	BTA11	HCD	APOB
Priones	BTA13	PRNP	PRNP
Tiroglobulina	BTA14	-	TG1
Diacylglicerol aciltransferasa	BTA14	-	DGAT1
Sindactilia	BTA15	Sindactilia	-
Maple syrup urine disease	BTA19	MSUD	BCKDHA
Pelo liso	BTA20	-	PRLR
Pelo liso	BTA23	-	PRL
Deficiencia de factor x1	BTA27	-	FX1
μ-calpaina	BTA29	-	CAPN1

^z Es el lugar donde se localiza un gen dentro del cromosoma. ^y Cualquier segmento de ADN encargado de una función específica en el organismo. Fuente: Casas y Kehrlí, (2016).

El reconocimiento de condiciones genéticas son un componente importante en la producción animal, dado que, varias de éstas condiciones genéticas pueden asociarse favorablemente con rendimiento y salud en bovinos, por ejemplo: el pelo liso en el ganado productor de leche bajo condiciones tropicales, debería ser de importancia económica si se introdujera, puesto que, al incrementar la productividad, el sistema de producción se haría más eficiente. Sin embargo, otras de las condiciones genéticas también pueden resultar perjudiciales, el caso más claro ocurre con: miostatina (MSTN), el gen responsable del doble musculo, este tipo de condición debe evaluarse cuidadosamente si se introduce en un hato, puesto que, al incrementar la masa muscular, existe mayor probabilidad de pérdidas económicas por problemas al parto. Casas y colaboradores (1999) indicaron que los animales con doble musculo (animales con dos copias del alelo MSTN) tienen hasta un 20% más de dificultad de parto que las razas sin doble musculo, no obstante, los animales con solo una copia del alelo MSTN aumentan la musculatura sin tener dificultad al parto.

A medida que los avances en genética molecular continúen, el impacto en la capacidad para detectar rápidamente genes relacionados a características de importancia económica, también mejorará.

2.4 Metodologías estadísticas

Meuwissen et al. (2001) propusieron un método que utiliza marcadores que cubren todo el genoma de modo que, toda la variación genética puede explicarse por los marcadores. En los estudios se supone que los paneles de marcadores son densos y cubren el genoma completo y están en desequilibrio de ligamiento con el QTL, por tanto, los marcadores se pueden utilizar para seleccionar el alelo favorable en cada QTL sin realmente identificar el QTL o el polimorfismo funcional. Estos autores utilizaron simulación y mostraron que el valor de cría de un individuo con solamente datos de marcadores se predijo con exactitud de 0.85, dada una población de referencia registrada para los marcadores y fenotipos.

El modelo paramétrico básico para evaluación genómica fue propuesto por Meuwissen et al. (2001):

$$y = \mu + \sum_{j=1}^p x_{ij}\beta_j + \varepsilon \quad o \quad y = \mu + X\beta + \varepsilon$$

donde $\varepsilon \sim N_n(0, \sigma^2)$, y es el vector de valores fenotípicos, X es la matriz diseño de orden $n \times p$ de los marcadores del genotipo, codificados como 0, 1, 2, para aa, Aa, AA, respectivamente, y β es el vector de parámetros desconocidos de los efectos de los marcadores.

La evaluación genómica de animales puede efectuarse en una o en dos etapas (Garrick, Taylor & Fernando, 2009). Cuando se realiza en dos etapas, en la primera se utiliza una población de entrenamiento (PEn) para obtener las ecuaciones de predicción, las cuales se usarán en una segunda etapa en la población de evaluación (PEv) para predecir los valores genómicos (VGn). En la PEv no se requieren los registros fenotípicos de los individuos, son necesarios sus genotipos (Alarcón-Zúñiga et al., 2015). Los animales utilizados en la PEv no deberán considerarse en la PEn. Los animales en ambas poblaciones deberán ser preferiblemente del mismo genotipo (Kachman et al., 2013). Los individuos de la PEv pueden evaluarse tan pronto como se pueda obtener una muestra de su tejido para realizar el genotipado.

Para predecir los valores genómicos (VGn), primero se obtienen ecuaciones de predicción basadas en efectos de los SNPs. El genoma completo se divide en segmentos pequeños, cuyos efectos se estiman en una población de referencia, donde los animales cuentan con fenotipos y genotipos. De ésta forma, los efectos de todos los *loci* que contribuyen a la variación genética son capturados, incluso si los efectos de los *locus* individuales son muy pequeños. En generaciones futuras, los animales pueden genotiparse para determinar qué segmentos de cromosoma llevan, y los efectos estimados de los segmentos heredados se realizan a continuación (Hayes et al., 2009).

Después del genotipado mediante un chip de SNPs, existe la posibilidad de que no se disponga de información para todos los genotipos. Esto puede resolverse mediante la imputación de genotipos. Con base en los genotipos conocidos de los animales, se reconoce el haplotipo que lleva el animal. Por tanto, el genotipo que falta puede leerse del genotipo de estos otros animales (Meuwissen, Hayes & Goddard, 2016). Los haplotipos son conjuntos de alelos en lugares cercanos que tienden a heredarse juntos.

Los programas estadísticos para la imputación de genotipos y para la estimación de los VGn (Genomic Estimated Breeding Value) son: Beagle (Browning & Browning, 2007), AlphaImpute (Hickey et al., 2012) e FImpute (Sargolzaei, Chesnais & Schekel, 2014).

En SG, a algunos animales no se les determina su genotipo, sin embargo, es necesario incluir su información fenotípica para la predicción de los VGn. Una forma de hacerlo es por SG de múltiples pasos o etapas: 1) se estiman pseudo-fenotipos para los animales genotipados; 2), la predicción genómica se realiza utilizando los pseudo-fenotipos y sus genotipos; y 3), el VG tradicional y VGn se combinan en un VG total (Meuwissen et al., 2016). La validación de VGn y la re-estimación de efectos, y de los parámetros de regularización se realizan a menudo después de que se agrega un lote de datos nuevo.

El número de marcadores moleculares (p) suele ser mayor que el número de observaciones (n) y, debido a esto, no es factible la estimación de los efectos de los marcadores a través de regresión múltiple por cuadrados mínimos ordinarios (OLS) (Pérez, de los Campos, Crossa & Gianola, 2010) o Máxima Verosimilitud (MV), debido a que la inversa de X^tX puede no existir, es muy probable que haya combinaciones lineales entre columnas o renglones de X y existen dependencias lineales conocidas como colinealidad o multicolinealidad (Nicolás, 2013), por ésta razón, se tienen que utilizar métodos de regularización.

Los programas para la predicción de VGn a escala grande incluyen GenSel (Garrick & Fernando, 2013), éste permite una variedad de métodos Bayesianos

de selección de variables; BLR (Pérez et al., 2010), implementa los métodos de Regresión Ridge Bayesiana y LASSO; BGLR (Pérez & de los Campos, 2014), utiliza varios modelos bayesianos paramétricos y procedimientos semi-paramétricos en un marco unificado; la familia BLUPF90 (Misztal, Legarra & Aguilar, 2009), que tiene el enfoque GBLUP en un sólo paso; MIX99 (Lidauer, Matilainen, Mäntysaari & Strandén, 2011) y ASReml4 (Gilmour, Gogel, Cullis & Thompson, 2009), que realiza algoritmos tipo REML para resolver ecuaciones de modelos mixtos.

Un elemento primario de la especificación del modelo en selección genómica es si los valores genéticos se aproximan mediante el uso de procedimientos de regresión lineal o utilizando métodos semi-paramétricos. Con las tecnologías de genotipado modernas, el número de marcadores, y el número de parámetros a estimar, puede superar ampliamente superior al número de registros. Para enfrentar a los problemas que surgen en las regresiones con p grandes y n pequeños, comúnmente, se usan procedimientos de estimación que realizan selección de variables, reducción de estimaciones o una combinación de ambas. Por tanto, un segundo elemento de elección del modelo pertenece al tipo de procedimiento de estimación de contracción utilizado.

A continuación, brevemente, se revisan algunos procedimientos de estimación de selección y reducción de las variables penalizadas y bayesianas más utilizados.

2.4.6 Métodos de regresión penalizada

Los métodos de regresión penalizada y bayesiana se basan en la minimización de la suma de cuadrados del error (SCE) sujeta a restricciones en los posibles valores de los estimadores para reducir su varianza, obteniéndose predicciones más precisas. Los métodos de estimación penalizados, como Regresión Ridge (Hoerl & Kennard, 1970), o procedimientos de reducción y selección de mínimos absolutos (LASSO, Tibshirani, 1996) pueden usarse para estimar los efectos de los marcadores (Pérez, de los Campos et al., 2010).

2.4.6.1 Regresión Ridge

Para predecir los VGn con el modelo de Regresión Ridge (RR-BLUP), todos los efectos de los SNPS se estiman simultáneamente en una población de referencia, la cual es genotipada y fenotipada; i. e., utilizando un modelo estadístico (asumiendo 30,000 SNPs):

$$y_i = \mu + X_{1i} \times \beta_1 + X_{2i} \times \beta_2 + \dots + X_{30000i} \times \beta_{30000} + e_i$$

donde y_i es el fenotipo del animal i ; μ es la media, X_{1i} es el genotipo del animal i para el marcador 1, β_1 es el efecto del marcador 1; y e_i es el residual.

Comúnmente, una suposición utilizada es que los efectos de SNP se distribuyen a priori normalmente con media 0 y varianza constante σ^2_ϵ . Este modelo considera los efectos de SNP como fijos. Los estimadores de los coeficientes de regresión se encuentran al solucionar las ecuaciones normales:

$$\hat{\beta} = (X^t X + I\lambda)^{-1} X^t y$$

donde λ es una constante para todos los SNPs y es conocido como el parámetro Ridge, el cual no tiene un valor específico. Algunas formas de determinar su valor son las siguientes (le Cessie & van Houwelingen, 1992; Lee & Silvapulle, 1988; Schaefer, 1994):

1. $\lambda_1 = \frac{1}{\beta^t \beta}$
2. $\lambda_2 = \frac{p+1}{\beta^t \beta}$
3. $\lambda_3 = \frac{(\lambda_1 - 100\lambda_p)}{99},$

donde λ_1 y λ_2 son el menor y mayor valor propio de $X^t X$.

El término λ , es el parámetro de penalización o parámetro Ridge, que controla la compensación entre la falta de ajuste y la complejidad del modelo, el cual permite superar el problema de multicolinealidad entre columnas de la matriz X que causa la no existencia de la matriz inversa de $X^t X$. Este estimador penalizado evita sobreajuste del modelo y estabiliza las estimaciones en comparación con la

regresión por mínimos cuadrados (de los Campos et al., 2013; Heslot, Yang, Sorrells & Jannink, 2012).

El valor genómico del candidato de selección i se predice como:

$$\widehat{VGn}_i = X_{1j} \times \hat{\beta}_1 + X_{2j} \times \hat{\beta}_2 + \dots + X_{30000j} \times \hat{\beta}_{30000}$$

donde $\hat{\beta}_1$ es el efecto estimado del SNP 1; y X_{1j} es el genotipo del animal i para el marcador SNP 1, y así sucesivamente.

2.4.6.2 Modelo GBLUP

En GBLUP, los VGn se predicen utilizando fenotipos y relaciones genómicas, que se basan en información de marcadores del genoma completo. El método GBLUP es, similar al tradicional BLUP, excepto que las relaciones de pedigrí se reemplazan por relaciones genómicas. Una ventaja práctica del enfoque de GBLUP es que todos los métodos y software tradicionales de BLUP pueden aplicarse: sólo se necesita reemplazar la matriz de pedigrí por las relaciones genómicas.

La relación de pedigrí entre dos hermanos completos es 0.5, lo que significa que dos hermanos completos representen 50% de sus alelos en común. Sin embargo, dos medios hermanos pueden compartir 60% o 40% de sus alelos, y ésta desviación de la expectativa basada en pedigrí del 50% se detecta por genotipificación; es aquí donde reside la mayor precisión de GBLUP comparada con el BLUP tradicional, porque las relaciones genómicas son más exactas que las relaciones basadas en pedigrí (Goddard, Hayes & Meuwissen, 2011; de los Campos et al., 2009). Esto último requiere que las estimaciones de la relación genómica se basen en un número grande de SNPs. Las relaciones entre razas son pequeñas y requieren un número de SNPs grande para utilizarse.

El modelo estadístico para GBLUP es:

$$y_i = \mu 1 + u_i + \varepsilon$$

La suposición es que $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$ y que, de forma independiente, $u \sim N(0, G\sigma_u^2)$, donde G es una matriz de relaciones genómicas derivada de marcadores, u_i es el valor de cría del animal i .

Los valores de cría de GBLUP y RR-BLUP son numéricamente equivalentes si u_i se define como:

$$u_i = X_{1i} \times \hat{\beta}_1 + X_{2i} \times \hat{\beta}_2 + \dots + X_{30000i} \times \hat{\beta}_{30000}$$

Los requisitos computacionales de GBLUP y RR-BLUP pueden ser muy diferentes. Por ejemplo, RR-BLUP requiere la estimación de 30,000 efectos SNP, y la solución simultánea de 30,000 ecuaciones, mientras que GBLUP requiere la estimación de n VGn y la solución de n ecuaciones, donde n es el número de animales. Como generalmente el número de animales genotipados es menor que 50,000, el método GBLUP tiene ventajas computacionales.

2.4.6.3 Reproducing Kernel Hilbert Space

El método Reproducing Kernel Hilbert Space (RKHS) utiliza una función *kernel* mediante la cual se transforma el conjunto de datos (marcadores) en un conjunto de distancias entre pares de observaciones, que da como resultado una matriz cuadrada que se utiliza como matriz diseño en un modelo lineal normal. Dado que la regresión no asume linealidad, se considera que los efectos no aditivos se podrían captar de mejor manera.

El modelo puede formularse como sigue (Heslot et al., 2012):

$$y = \mu 1 + u_i + \varepsilon$$

donde:

μ es el vector de efectos fijos. La suposición es que $\varepsilon \sim (0, \sigma_\varepsilon^2 I)$ y que, de forma independiente, $u \sim N(0, K_h \sigma_u^2)$. $K_h = \{X_i, X_{i'}\}$ es una matriz de incidencias, que depende de la elección del kernel y del parámetro h , que de forma semejante a una matriz de correlaciones mide la similitud genómica entre los genotipos.

2.4.7 Alfabeto Bayesiano

La información previa en RR-BLUP y GBLUP supone que los efectos SNP se distribuyen normalmente con la misma varianza para cada SNP. Biológicamente, se puede esperar que algunos SNPs, que están cerca de un gen, tengan un efecto y muchos otros no. Esto indica que, las estimaciones de los efectos de los

marcadores se penalizan en la misma medida, y puede no ser apropiado si algunos marcadores están localizados en regiones no asociadas con la varianza genética, mientras que otros están vinculados a QTLs (Goddard & Hayes, 2007).

En muchos casos se puede tener información previa adicional sobre los marcadores y tal vez se quiera incorporar dicha información. Recientemente varios métodos que incorporan información previa que supone que una fracción π de los SNP tiene un efecto y una fracción $(1-\pi)$ no lo tiene en absoluto se han desarrollado.

El modelo utilizado para estos métodos es:

$$y_i = \mu + I_1 \times X_{1i} \times b_1 + I_2 \times X_{2i} \times b_2 + \dots + I_{30000} \times X_{30000i} \times b_{30000} + e_i$$

donde:

I_i es una variable indicadora con valores 0 o 1 que indica si SNP _{i} tiene un efecto o no.

Los métodos "bayesianos" se llaman así, debido a que, usan una distribución previa de los efectos de SNP, tales como: BayesA y BayesB de Meuwissen et al. (2001) o el LASSO Bayesiano de Park y Casella (2008). La distribución previa de los efectos SNP utilizados por los métodos bayesianos tiene más sentido biológico que si se supone que todos los SNP tienen un efecto y que todos los efectos son muy pequeños. Una de las ventajas grandes de los modelos bayesianos es que, potencialmente, se puede incluir diferentes tipos de información previa en el modelo (de los Campos et al., 2013), por ejemplo:

- la ubicación del marcador en el genoma,
- si el marcador está situado en una región codificante o no codificante,
- si el marcador está en una región que alberga genes que se cree afectan la característica de interés, y
- cualquier información previa sobre el marcador de un estudio independiente, como *p-values* o estimaciones de efectos derivados de un estudio de asociación a nivel genómico.

En estudios de simulación, los métodos no lineales superan a GBLUP (Meuwissen & Goddard, 2010), con datos reales, los métodos no lineales son

superiores para algunas características (Erbe et al., 2012). Esto puede explicarse por lo siguiente: 1) Existen muchos genes que afectan las características económicamente importantes, de modo que suponiendo que todos los SNPs tienen un efecto es aproximadamente cierto; 2) el desequilibrio de ligamiento (la asociación no aleatoria entre dos *loci*) se extiende a distancias genómicas grandes en las poblaciones de ganado, de manera que muchos SNPs están asociados con un gen; y 3) la densidad del panel de marcadores SNP no es lo suficientemente grande, por lo que tantos SNPs son necesarios para explicar conjuntamente el efecto QTL. En consecuencia, la superioridad de los métodos no lineales es más clara cuando el chip SNP de 50k se reemplaza por uno de densidad alta, 777k, y cuando los datos incluyen razas múltiples (Brøndum et al., 2015).

Kemper et al. (2015) indicaron que la exactitud de las predicciones genómicas obtenidas con BayesR aumentó 8% para el bovino lechero australiano de una población de referencia Holstein/Jersey. Además, demostraron que éste método no lineal, panel de marcadores de alta densidad, combinado con una población de referencia multirracial que incluyó vacas y toros, resultó en predicciones genómicas mayores, especialmente para una raza que no se incluyó en la población de referencia.

2.4.8 Validación cruzada

La validación cruzada (VC) se utiliza para estimar el poder predictivo de modelos diferentes. Las formas de utilizar VC son: pruebas de entrenamiento únicas o replicadas y VC replicada. La diferencia principal es que en la segunda todos los individuos están en la población de entrenamiento al menos una vez, mientras que, en la primera, algunos individuos nunca son parte de la población de entrenamiento. La partición de las poblaciones de entrenamiento y validación afectará la precisión alcanzada. La exactitud de la predicción se puede utilizar para evaluar el modelo subyacente y comparar modelos alternativos (Daetwyler et al., 2013). Si la población de validación está más relacionada con la población

de entrenamiento que los candidatos de selección, entonces la estimación de precisión de predicción aumentará.

En forma general, la VC consiste en dividir la información aleatoriamente en conjuntos disjuntos (por ejemplo 10). De ésta manera, los datos quedarán divididos en 10 conjuntos de tamaño aproximado $k=[n/10]$.

Por ejemplo, si n es múltiplo de 10:

$$y^* = (y_{[1]}, \dots, y_{[k]}, y_{[k+1]}, \dots, y_{[2k]}, \dots, y_{[9k+1]}, \dots, y_{[10k]})^T$$

donde:

$y_{[i]}$ es cualquier elemento de y que fue ordenado para que ocupe la posición $[i]$.

La matriz X también se ordena de tal manera que sus n hileras correspondan a los n elementos de y^* , resultando X^* .

El método procede de la siguiente manera:

1. Se ajusta el modelo $y_{-1} = X_{-1}\beta$ y se obtienen los valores $\hat{\beta}_{-1}$, donde y_{-1} es el vector y^* de donde se suprimieron las k observaciones correspondientes al conjunto 1 y X_{-1} es la matriz X^* de la cual se eliminaron las hileras correspondientes al conjunto 1.
2. Con el vector de coeficientes estimados $\hat{\beta}_{-1}$, se predicen los k valores eliminados:

$$\hat{y}_{-1} = X_{-1}\hat{\beta}_{-1} + \sigma_\varepsilon z,$$

donde $z \sim N(0, 1)$, $\hat{y}_{-1} = (y_{[1]}, \dots, y_{[k]})^t$ y X_{-1} es la matriz que sólo contiene las k hileras correspondientes al conjunto 1.

3. Se repiten los pasos 1 y 2 para cada uno de los conjuntos restantes.
4. Cuando se tienen todas las observaciones predichas de los 10 conjuntos:

$$\hat{y}^* = (y_{[1]}, \dots, y_{[k]}, y_{[k+1]}, \dots, y_{[2k]}, \dots, y_{[9k+1]}, \dots, y_{[10k]})^t,$$

se estima su correlación con el vector para predecir valores futuros y^* .

La correlación entre los valores reales y los predichos en el paso 4 es de interés para conocer el modelo con mejor capacidad predictiva, entre más grande sea esta correlación, mejor será el modelo utilizado para predecir valores futuros.

Usualmente n no es múltiplo de 10, en cuyo caso algunos de los grupos tendrán más observaciones que el resto, pero la idea general es la misma (López, 2012).

Goddard (2009) presentó un método determinístico para calcular la precisión de los VGn cuando la ecuación de predicción se estima en una población de referencia, y para un determinado nivel de heredabilidad (Figura 2).

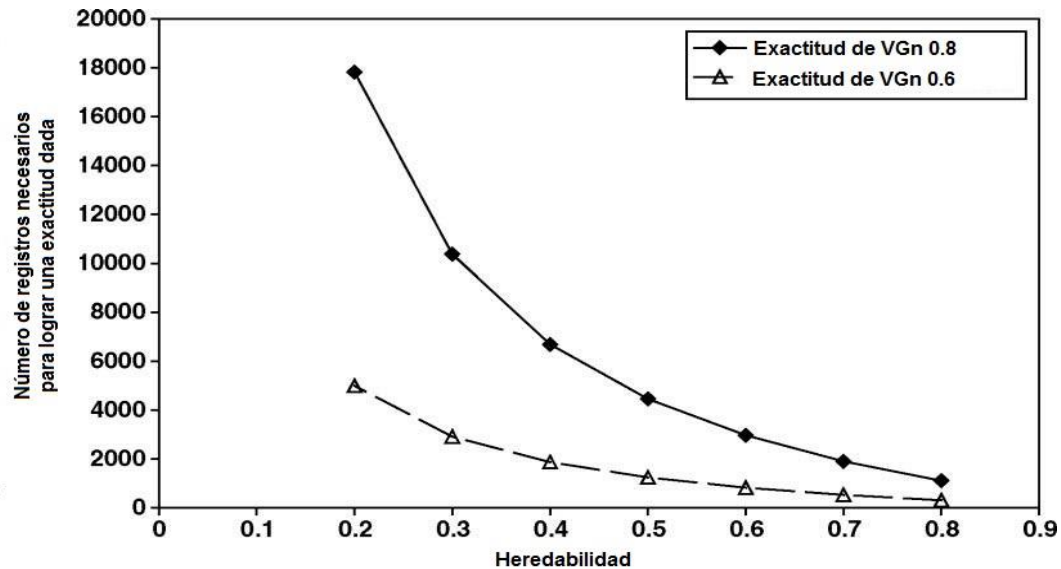


Figura 2. Número de registros fenotípicos necesarios para lograr una exactitud determinada del valor genético predicho, dada la heredabilidad de la característica. Se supone un tamaño efectivo de la población (N_e) = 1,000 y una distribución normal de los efectos de QTLs. Fuente: Hayes et al. (2009).

Por ejemplo: para características de baja heredabilidad, se requerirá una gran cantidad de registros en la población de referencia para lograr una alta exactitud de VGn en animales no genotificados.

Erbe y colaboradores (2010) estudiaron los cambios en los resultados al utilizar diferentes números de animales para la población de referencia y de validación para encontrar estrategias de subdivisión óptimas para los diferentes objetivos de la VC en animales Holstein. Estos autores observaron las correlaciones más altas cuando el número de animales en el conjunto de validación (n_v) fue pequeño ($n_v = 100$ y $r = 0.625$ vs $n_v = 1500$ y $r = 0.536$). Una subdivisión quíntuple, que utiliza 20% de los datos como conjunto de validación, parece ser aceptable.

2.5 Herramientas basadas en fenotipos y pedigrí versus asistidas por marcadores en la predicción de valores genéticos

En los esquemas tradicionales de pruebas de progenie, un toro utilizado para IA tendría su primera evaluación genética confiable hasta que un número mínimo de su progenie complete registros de rendimiento (mínimo 64 meses), mientras que, al usar ecuaciones de predicción genómica, un ternero puede tener una evaluación genética confiable tan pronto como se pueda genotipar. Por tanto, la selección puede ocurrir temprano en la vida del animal, reduciendo en gran medida el intervalo generacional y los costos de prueba (Dürr & Philipsson, 2014). En la Figura 3 se presenta un cronograma de prueba de progenie tradicional.

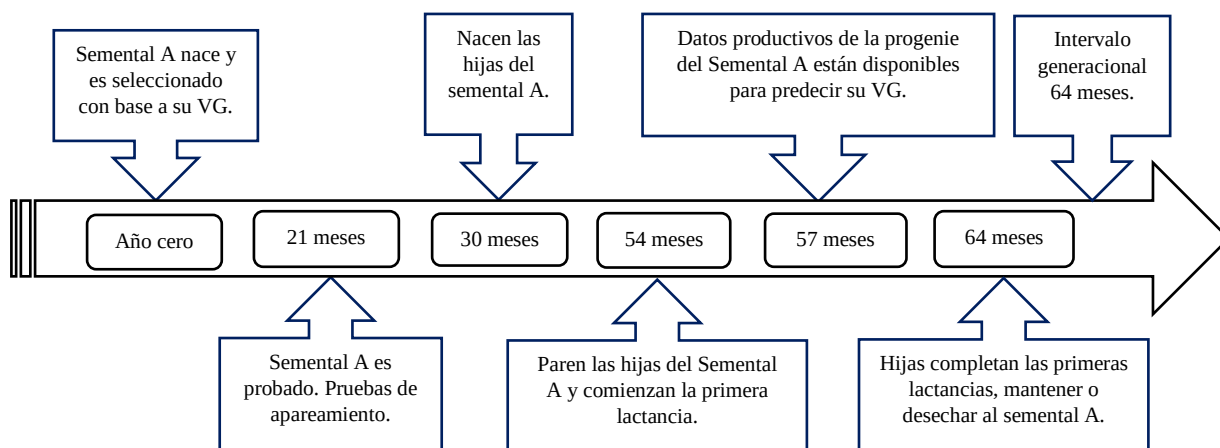


Figura 3. Intervalo entre generaciones en bovinos lecheros en un programa de prueba de progenie tradicional.

Con SG en lugar de esperar 4.5 años para utilizar toros probados por su progenie, ahora las compañías utilizan los mejores toros jóvenes probados por ADN en aproximadamente un año de edad, lo que reduce notablemente el intervalo generacional.

La Figura 4 ilustra la línea de tiempo para un programa de inseminación artificial basado en evaluación genómica. Por ejemplo: el semental B tiene 54 meses cuando los fenotipos de sus hijas están disponibles para el cálculo de su VG y los criadores ya están utilizando semen de sus mejores nietos para producir sus bisnietos y bisnietas. Antes de la introducción de la SG, los genetistas,

generalmente, esperaban hasta que las vacas tuvieran al menos 2 años de edad antes de usar tecnologías reproductivas, como la ovulación múltiple y la transferencia de embriones. Sin embargo, las pruebas de ADN de becerras proporcionan VGn precisos para permitir que decisiones de selección se realicen antes que las vaquillas alcanzan la madurez sexual (Schaeffer, 2006; Schefers & Weigel, 2012).

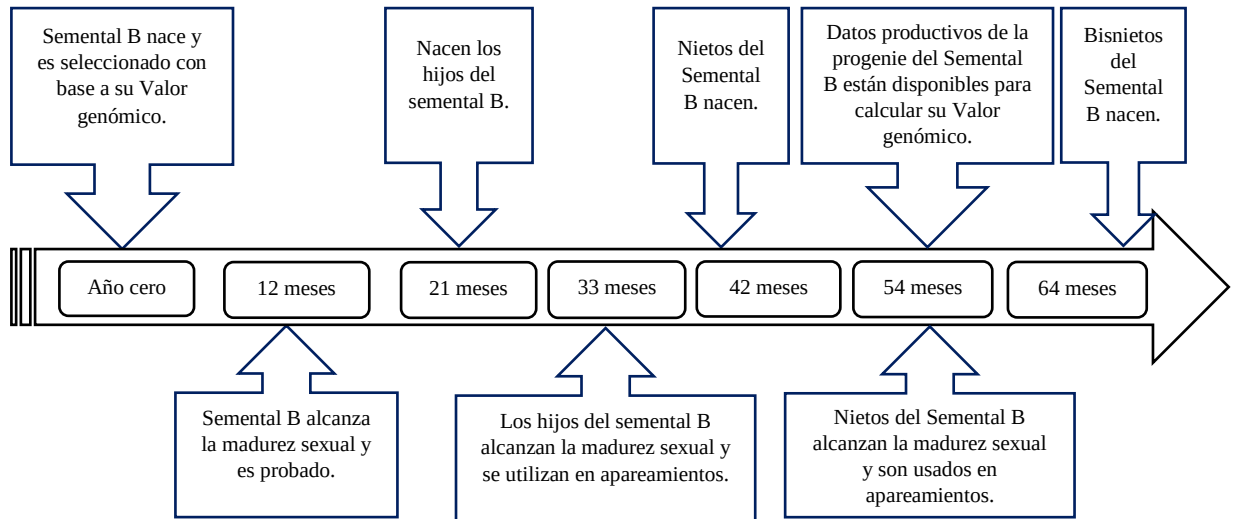


Figura 4. Cronograma de un programa de cría que utiliza inseminación artificial basado en evaluación genómica.

En los estudios de simulación (Meuwissen et al., 2001; Habier et al., 2007) se demostró una mayor exactitud de las predicciones mediante evaluación genómica con respecto a las basadas en fenotipos y pedigrí, independientemente de la característica o modelo de elección. En general, los estudios han confirmado la superioridad de SG relativa a las predicciones basadas en pedigrí. El caso más claro ocurre en bovinos Holstein-Friesian. El potencial de SG también se ha confirmado en varias razas de bovinos productores de carne (Garrick 2011; Saatchi et al., 2011), ovinos (Daetwyler et al., 2012), pollos (Gonzalez-Recio et al., 2008), gallinas (Wolc et al., 2011a,b) y varias especies de plantas (Heslot et al., 2012).

Meuwissen et al. (2001) compararon el método BLUP y análisis Bayesianos en la exactitud de las predicciones de VGn en individuos genotipados y sin registros productivos. Estos autores encontraron que los métodos bayesianos incrementaron la exactitud de las predicciones hasta 0.85 comparada con BLUP donde la exactitud fue 0.73. El método BayesA resultó con 9% más de exactitud que BLUP y la exactitud de BayesB excedió 16% a la de BLUP. También indicaron que las exactitudes de estos métodos son altas para esquemas de selección donde los animales no tienen información fenotípica propia.

Por otro lado, Calus et al. (2014) compararon un modelo convencional con datos de pedigrí vs modelos de predicción genómica (GBLUP, RR-BLUP y BayesC) en la exactitud de las predicciones para número de huevos en el primer periodo de postura para gallinas. Los modelos de predicción genómica arrojaron una precisión de 0.13 a 0.16 más alta que BLUP basado en pedigrí.

Daetwyler et al. (2012) registraron ganancias en la precisión de VGn entre 0.07 y 0.31 para características de la canal y calidad de la carne en ovejas australianas en comparación con una exactitud de 0.09 para BLUP basado en pedigrí.

La superioridad de SG, con relación a las predicciones basadas en pedigrí, no siempre ha sido tan alta como se anticipó por simulación. Muchos estudios han demostrado que factores como tamaño del conjunto de datos de referencia (Meuwissen et al., 2001; VanRaden & Sullivan 2010), heredabilidad de las características, número de *loci* que afectan las características (Daetwyler et al., 2008), distancia entre marcadores, tamaño de la población de entrenamiento, número de animales con información genómica y fenotípica (Hayes et al., 2009) y grado de las relaciones genéticas entre la población de entrenamiento y la población de validación (Habier et al., 2007) pueden afectar la exactitud de la predicción de valores genéticos.

En algunos estudios se han demostrado los beneficios de extender los modelos de SG mediante la adición de la información de pedigrí (de los Campos et al., 2009; Crossa et al., 2010). Sin embargo, los beneficios de modelar

conjuntamente los datos de pedigrí y los marcadores en relación con un modelo de solamente de marcadores parecen desaparecer a medida que la densidad de marcadores aumenta (Vazquez et al., 2010). Ahora se acepta que la mejor manera de utilizar la información genómica en programas de mejoramiento genético es permitir que la información genómica contribuya a la evaluación genética de los animales, los VG tradicionales basados en fenotipo y pedigrí se combinan con el VG genómico o la predicción genética basada sólo en marcadores. Este enfoque considera la ponderación de las dos fuentes de información que es una función de la precisión de las dos predicciones (Miller, 2010).

2.6 Progreso genético utilizando selección genómica

Con SG, algunos criadores ahora pueden obtener información temprana sobre los valores genéticos de sus animales, lo que cambia sus oportunidades de comercialización. Por otra parte, los productores comerciales podrán beneficiarse del progreso genético logrado por los criadores de ganado de registro, adquiriendo sementales y vientres con calidad genética certificada, para mejorar con mayor certeza la productividad de sus hatos.

Actualmente, la SG es el principal reto en mejoramiento genético animal, se ha visto que su aplicación puede incrementar la tasa de progreso genético. Sin embargo, el éxito de su implementación depende de cada esquema de selección. Por ejemplo, Wiggans et al. (2017) indicaron que, a partir de enero de 2009 estuvieron disponibles en USA las primeras evaluaciones genómicas oficiales para Holstein-Friesian y Jersey, en agosto de 2009 para Brown Swiss, en abril de 2013 para Ayrshire, y en abril de 2016 para Guernsey. El genotipado inicial y el procesamiento de muestras comerciales fue realizado por GeneSeek (Lincoln, Nebraska, ahora parte de Neogen), Genetics Visions (Middleton, Wisconsin) y Zoetis Genetics (Kalamazoo, Michigan), con reportes de exactitud mayor que 0.8 para características de producción y 0.7 para fertilidad, longevidad y conteo de células somáticas (Wiggans, VanRaden & Coopel, 2011; Lund et al., 2011). Además, se están evaluando genómicamente otras características en las razas

Ayrshyre, Suizo Americano, Guernsey, Holstein y Jersey incluyendo tasa de concepción en primerizas y adultas, estatura, profundidad del cuerpo, ángulo de la grupa, vista de aplomos (lateral y posterior), ángulo de la pezuña, altura y profundidad de la ubre, posición de pezones anteriores y largo del pezón (Wiggans et al., 2017). Estos reportes en progreso genético se han logrado gracias la disponibilidad de chips de costo bajo, lo que ha provocado año con año el incremento en el número de genotipados en poblaciones de animales en EE. UU. A. (Figura 5). Además, los SNP están reemplazando gradualmente los micro-satélites para la verificación de parentesco.

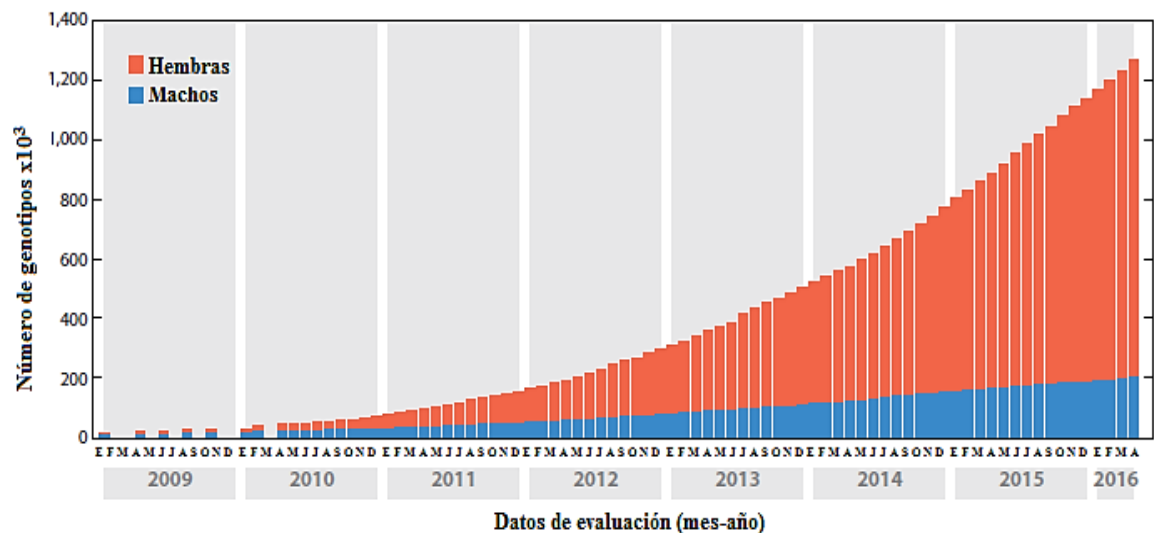


Figura 5. Número de animales genotipados e incluidos en las evaluaciones genómicas para bovinos lecheros en los EE. UU. desde enero de 2009.

En bovinos productores de carne, el desarrollo e implementación de las evaluaciones genómicas se han visto obstaculizados por diversos factores, incluyendo:

➤ Múltiples razas y sus cruzas

Los genotipos utilizados en producción bovina para carne son cruza especializadas en la producción de carne, de modo que, existe una diversidad grande de razas y cuando se desea realizar evaluaciones genómicas, las razas pueden o no estar suficientemente representadas en la población de referencia (Kachman et al., 2013). Las predicciones genómicas para los animales

candidatos de una raza pueden correlacionarse negativamente con su rendimiento fenotípico, resultante si esa raza no está suficientemente representada en la población de referencia (Kachman et al., 2013). De manera similar, las predicciones genómicas para una raza utilizando una población de referencia que representa una selección de países pueden estar correlacionadas negativamente con el rendimiento de animales de la misma raza procedentes de un país diferente (Saatchi, Ward & Garrick, 2013).

➤ Menos uso de inseminación artificial

Una proporción grande de bovinos lecheros, p. ej. Holstein, dependen de pocos toros de inseminación artificial (IA). Por el contrario, una menor proporción de animales en bovinos para carne se generan a partir de IA. Menos uso de IA generalmente implica menos toros con evaluaciones genéticas de alta precisión. Colectivamente, éstas restricciones requieren una población de referencia grande, difícil y costosa de montar. El nivel bajo de uso de IA en bovinos carne también contribuye a una conectividad genética deficiente.

➤ Bajos niveles de fenotipado

Cuanto mayor es la fiabilidad de la información fenotípica en la población de referencia, la precisión de las predicciones genómicas es mayor (Daetwyler et al., 2008; Garrick et al., 2009). Las estrategias de fenotipado en los sistemas de producción de carne tienden a ser pobres, especialmente, en poblaciones comerciales y para características medidas tarde en la vida de los animales.

En el resto de las especies de producción animal; están disponibles paneles de 40 a 65k SNPs, incluyendo cerdos, aves, ovinos y equinos. Recientemente, los paneles con más de 700k SNPs están disponibles para bovinos y paneles de densidad mayor también se están desarrollando en otras especies (Dekkers, 2012). El costo de las pruebas va de \$1,400.00 a \$1,700.00 pesos para chips de densidad alta, y de \$800.00 a \$1,100.00 pesos para chips de densidad baja (Dekkers, 2012). Se espera que los costos se reduzcan con el tiempo.

La secuenciación del genoma en aves fue el primero en estar disponible en 2004. Desde entonces, el desarrollo de paneles de SNP para aves ha sido rápido. Tres

paneles de SNP de tamaños 6, 12 y 42k están ahora disponibles (Ibañez-Escriche & González-Recio, 2011).

El chip de marcadores SNP para ovinos estuvo disponible en 2009. El esquema de registro de datos para éstas especies es limitado y, por tanto, la ganancia genética obtenida con la estrategia de SG puede ser menor que la obtenida con selección tradicional. Además, el costo del genotipado en relación con el valor del animal sigue siendo una limitante económica para adoptar ésta nueva tecnología en la cría de ovejas y cabras.

En porcinos, las ganancias adicionales por reducción del intervalo generacional son limitadas, aunque todavía se puede lograr disminuir aproximadamente 25% en el intervalo generacional con la introducción de SG (Meuwissen et al., 2016).

En conejos, tres factores principales limitan la aplicación de SG: 1) el chip no está disponible, aunque el genoma del conejo está secuenciándose, 2) los conejos tienen un intervalo de generación corto, y 3) el valor económico por animal de ésta especie es bajo en comparación con el costo de genotipado. Por lo anterior, es difícil justificar el costo económico de SG en la especie.

Actualmente, no se dispone de mapas densos ni de plataformas de genotipado para peces. Por tanto, la aplicación real de SG en peces no se conoce, sólo se han realizado algunos estudios basados en simulaciones por computadora.

2.7 Perspectivas de selección genómica

La tendencia clara de la SG en los próximos años, es que, el genotipado (especialmente con plataformas de genotipificación de densidad baja) va a utilizarse ampliamente, y es probable que una mayor proporción de animales reproductores se genotipificarán en el futuro. Sin embargo, existe el potencial de combinar paneles de densidad baja con paneles >10,000 SNP, para inferir los genotipos que se obtendrían con paneles más grandes, 500k, o incluso la secuencia completa donde esté disponible. Lo anterior, es posible con algoritmos de imputación efectivos (Meuwissen et al., 2016).

En bovinos, esto podría significar genotipificar todos los toros con un panel de densidad alta y, luego, la progenie podría tener todos sus genotipos de densidad alta atribuidos a un panel de densidad menor y más rentable (Miller, 2010). Ésta perspectiva de imputar genotipos es prometedora. Al respecto, Sargolzaei et al. (2014) indican que cuando el panel de SNPs tiene una densidad alta que 3000 marcadores, la imputación es muy exitosa con precisión hasta 0.99. La SG ofrece dos oportunidades no utilizadas plenamente hasta ahora: 1) en la combinación de tecnologías reproductivas se podría disminuir el intervalo generacional, y en combinación con la ovulación múltiple y la transferencia de embriones, podría utilizarse para coleccionar embriones de calidad suprema para producir la generación de animales del futuro, 2) la ecuación de predicción se podría evaluar en animales comerciales, en lugar de animales élite. Por ejemplo, los animales comerciales son a menudo cruzados, se crían en ambientes adversos, y se puede recolectar información de características que no se miden a nivel de estudios, como calidad de la carne y resistencia a las enfermedades (Berry, García & Garrick, 2016; Meuwissen et al., 2016; Simianer, 2016).

Finalmente, la metabolómica integrada con SG podría traer grandes beneficios. Ésta disciplina identifica, cuantifica y caracteriza un número grande de metabolitos en leche, plasma y suero, proporcionando una interesante comprensión de los denominados fenotipos intermedios entre el nivel genómico y los fenotipos finales o externos, que en bovinos podrían ser características de producción como tasa de crecimiento, producción de leche, deposición de grasa y trastornos metabólicos y reproductivos (Fontanesi, 2016). La integración de la metabolómica con la genómica del bovino ha sido presentada en pocos estudios con perspectivas prometedoras.

La información genómica puede utilizarse para diseñar estrategias de cruzamiento más eficientes, para evaluar y conservar la diversidad genética y para manejar la producción. En México existen programas nacionales de evaluaciones genéticas periódicas en poblaciones de bovinos productores de carne. Sin embargo, no se han reportado estudios relacionados con el uso de

información genómica utilizando SNPs. Por lo que la implementación de evaluaciones genómicas o que combinen resultados de las dos áreas de estudio representa una oportunidad.

2.8 Literatura Citada

- Alarcón-Zuñiga, B., Ramírez-Flores, F., Ruíz-Flores, A., Ramírez-Valverde, R., Saavedra-Jiménez, L. A., & Zepeda-Batista, J. L. (2016). Comparison of the accuracy of genomic values of animals predicted through analysis with two alternative models. *Agrociencia*, 49, 613-622.
- Avice, J. C. (2004). *Molecular Markers, Natural History, and Evolution*. 2da Edición. Ediciones Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts, 541 p.
- Berry, D. P., García, J. F., & Garrick, D. J. (2016). Development and implementation of genomic predictions in beef cattle. *Animal Frontiers*, 6(1), 32-38, doi:10.2527/af.2016-0005.
- Brøndum, R. F., Su, G., Janss, L., Sahana, G., Guldabrandtsen, B., Boichard, D., & Lund, M. S. (2015). Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *Journal of Dairy Science*, 98(4), doi:10.3168/jds.2014-9005.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, 81, 1084-1097, doi:10.1086/521987.
- Calus, M. P., Huang, H., Vereijken, A., Visscher, J., Napel, J., & Windig, J. (2014). Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genetics Selection Evolution*, 1, 46-47.
- Casas, E., & Kehrli, M. E. (2016). A review of selected genes with known effects on performance and health of cattle. *Frontiers in Veterinary Science*, 3, 1-11, doi:10.3389/fvets.2016.00113.
- Casas, E., Keele, J. W., Fahrenkrug, S. C., Smith, T. P., Cundiff, L. V., Stone, R. T. (1999). Quantitative analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating an inactive myostatin allele. *Journal Animal Science*, 77, 1686-92, doi:10.2527/1999.7771686x.
- Crossa, J., de los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J., . . . Braun, H. J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186, 713-724, doi:10.1534/genetics.110.118521.
- Cuevas P., R. G. (2013). *Uso de marcadores moleculares para verificar la genealogía de bovinos jersey*. (Tesis de Maestría en Ciencias, Universidad Autónoma Chapingo, Chapingo, México.) Consultada en https://chapingo.mx/produccionanimal/administrador/components/com_jsearch/files/theses/PPA_MC_046_11_13_RGP-RGCP.pdf
- Daetwyler, H. D., Calus, M., Pong-Wong, R., de los Campos, G., & Hickey, J. (2013). Genomic prediction in animals and plants: simulation of data,

- validation, reporting and benchmarking. *Genetics*, 193, 347-365, doi:10.1534/genetics.112.147983.
- Daetwyler, H. D., Hickey, J. M., Henshall, J. M., Dominik, S., & Gredler, B. (2010). Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science*, 2010, 1004-1010, doi:10.1071/AN10096.
- Daetwyler, H. D., Villanueva, B., & Woolliams, J. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *Plos One*, 3(10), 1-8, doi:10.1071/AN10096.
- Daetwyler, H., Swan, A., HJ, J., Van der Werf, J., & Hayes, B. (2012). Accuracy of pedigree and genomic predictions multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution*, 44(1), 1-11.
- de los Campos, G., Naya, H., Gianola, D., Crossa J., Legarra A., Manfredi, E., Weigel, K., & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182, 375-385, doi: 10.1534/genetics.109.101501.
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193, 327-345, doi:10.1534/genetics.112.143313.
- Dekkers, M. J. (2012). Application of genomics tools to animal breeding. *Current Genomics*, 13, 207-212.
- Dürr, J., & Philipsson, J. (2014). International cooperation: The pathway for cattle genomics. *Animal Frontiers*, 2(1), 16-21, doi:10.2527/af.2011-0026.
- Elzo, M. A. (2014). Repeatability Models. Animal breeding notes. En *Repeatability Models. Animal breeding notes* (pág. Chapter 15). University of Florida.
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., . . . Goddard, M. E. (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, doi:10.3168/jds.2011-5019.
- Erbe, M., Pimentel, E., Sharifi, A., & Simianer, H. (2010). Assessment of cross-validation strategies for genomic prediction in cattle. *Genetics*.
- Fontanesi, L. (2016). Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Animal Frontiers*, 6(1), 73-79, doi:10.2527/af.2016-0011.
- Garrick, D. (2011). The nature, scope and impact of genomic prediction in beef cattle in the United States. *Genetics Selection Evolution*, 43, 17, doi:10.1186/1297-9686-43-17.

- Garrick, D. J., & Fernando, R. L. (2013). Implementation a QTL detection study (GWAS) using genomic prediction methodology. *Methods in Molecular Biology*, 1019, 275-298, doi:10.1007/978-1-62703-447-0_11.
- Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41(55), doi:10.1186/1297-9686-41-55.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). *ASReml User Guide Release 3.0*.
- Goddard, M. E. (2009). Genomic selection: Prediction of accuracy and maximization of long term response. *Genetics*, 136, 245–257, doi:10.1007/s10709-008-9308-0.
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124, 323-330.
- Goddard, M. E., Hayes, B. J., & Meuwissen T. H. E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal Animal Breeding Genetics*, 128, 409-421. doi: 10.1111/j.1439-0388.2011.00964.
- Gonzalez-Recio, O., Gianola, D., Long, N., Weigel, K. A., & Rosa, G. J. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, 178, 2305-2313, doi:10.1534/genetics.107.084293.
- Habier, D., Fernando, R. L., & Dekkers, J. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177, 2389-2397, doi:10.1534/genetics.107.081190.
- Hannachi , M., & Tichit , M. (2016). Does biotechnological innovation require organizational innovation? learning from the cattle breeding industry in France. *Animal Frontiers*, 6(1), 80-85, doi:10.2527/af.2016-0012.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *Journal Dairy Science*, 92(2), 433-443, doi:10.3168/jds.2008-1646.
- Henderson, C. R. (1950). Estimation of genetic parameters. *Ann. Math. Stat.* 21, 309.
- Henderson, C. R. (1984). Best Linear Unbiased Prediction of performance and breeding value. Cornell University and the University of Illinois.
- Heslot, N., Yang, H.-P., Sorrells, M. E., & Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Science Society of America*, 52, 146-160, doi:10.2135/cropsci2011.06.0297.
- Hickey, J. M., Kinghorn, B. P., Tier, B., Clark, S. A., Van de Werf, J. H., & Gorjanc, G. (2012). Genomic evaluations using similarity between haplotypes,.

- Journal of Animal Breeding and Genetics*, 130(4), 1-11, doi:10.1111/jbg.12020.
- Hoerl, A. E., & Kennard, R. B. (1970). Ridge regression: biased estimation for nonorthogonal problems. *American Statistical Association*, 12(1), 5-67.
- Ibañez-Escriche, N., & Gonzalez-Recio, O. (2011). Review. Promises, pitfalls and challenges of genomic selection in breeding programs. *Spanish Journal of Agricultural Research*, 9(2), 404-413.
- Interbull. (2017). MACE explained. Available at; <http://www.interbull.org/ib/interbullactivities>. Consulted on February 12th, 2018.
- Kachman, S. D., Spangler, M. L., Bennett, G. L., Hanford, K. J., Kuehn, L. A., Snelling, M. W., . . . Pollak, E. J. (2013). Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution*, 45(30), doi:10.1186/1297-9686-45-30.
- Kemper, K., Reich, C., Bowman, P., Vander Jagt, C., Chamberlain, A., Mason, B., . . . Goddard, M. (2015). Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution*, 47, 29, doi:10.1186/s12711-014-0074-4.
- le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Appl. Statist*, 41(1), 191-201.
- Lee, A. H., & Silvapulle, M. J. (1988). Ridge estimation in logistic regression. *Comm. in Statistics-Theory and Methods*, 17(4), 1231-1257
- Lidauer, M. H., Matilainen, K., Mäntysaari, E. A., & Strandén, I. (2011). *MTT Agrifood Research Finland; Technical reference guide for MiX99*. Obtenido de <http://www.mtt.fi/BGE/Software/MiX99>.
- López C., M. A. (2012). Aplicación del Elastic Net LASSO y modelos relacionados en Selección Genómica basados en Marcadores Moleculares. (Tesis de Maestría en Ciencias, Colegio de Postgraduados, Montecillo, Estado de México). Disponible en http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/712/Lopez_Cruz_MA_MC_Estadistica_2012.pdf?sequence=1&isAllowed=y. Consultado el 16 de enero de 2018.
- Lund, M. S., PW de Roos, A., G de Vries, A., Druet, T., Ducrocq, V., Fritz, S., . . . Su, G. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genetics Selection Evolution*, 43, 43, doi:10.1186/1297-9686-43-43.
- Martínez, M. C., Helguera, M., & Carrera, A. (2011). Marcadores Moleculares. en: *Biotecnología y Mejoramiento Vegetal II*. Levitus, G., Echenique, V., Rubinstein, C., Hopp, E., & Mrogunski L. (Eds.). 2da Edición. Ediciones Instituto Nacional de Tecnología Agropecuaria. Buenos Aires, Arg. 648 p.

- Martínez-Niño, C., Manrique-Perdomo, C., & Elzo, M. (2012). Cattle genetic evaluation: a historical perception. *Revista Colombiana de Ciencias Pecuarias*, 25, 293-311.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6(1), 6-14, doi:10.2527/af.2016-0002.
- Meuwissen, T., & Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing, 185, 623-631. *Genetics*, 185, 623-631, doi:10.1534/genetics.110.116590.
- Miller, S. (2010). Genetic improvement of cattle through opportunities in genomics. *Revista Brasileira de Zootecnia*, 39, 247-255.
- Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92, 4648-4655.
- Misztal, I., Legarra, A., & Aguilar, I. (2014). Using recursion to compute the inverse of the genomic relationship matrix. *Journal of Dairy Science*, 97, 3943-3952.
- Mrode, R. A., & Thompson, R. (2005). *Linear Models for the Prediction of Animal Breeding Values*. 2nd Edition. Cronwell Press, Trowbridge, UK.
- NCBI. (2017). Bos taurus genome. Disponible en: <https://www.ncbi.nlm.nih.gov/genome/?term=BOS%20TAURUS%20>. Consultado el 16 de enero de 2018.
- Nicolás P., E. (2013). Un método de regresión bayesiana para selección genómica. (Tesis de Maestría en Ciencias, Colegio de Postgraduados, Montecillo, Texcoco, Estado de México). Disponible en: http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2089/Nicolas_Popoca_%20E_MC_Estadistica_2013.pdf?sequence=1. Consultado el 15 de noviembre de 2017.
- Núñez-Domínguez, R., Ramírez-Valverde, R., Ruíz-Flores, A., & Domínguez-Viveros, J. (2003). *Evaluación genética se sementales Suizo Europeo*. Universidad Autónoma Chapingo, Departamento de Zootecnia, Chapingo, México.
- Núñez-Domínguez, R., Ramírez-Valverde, R., Ruíz-Flores, A., & Domínguez-Viveros, J. (2014). *Evaluación genética se sementales Suizo Europeo*. Universidad Autónoma Chapingo, Departamento de Zootecnia, Chapingo, México.

- Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681-686, doi:10.1198/016214508000000337.
- Pérez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2), 483-495, doi:10.1534/genetics.114.164442.
- Pérez, P., de los Campos, G., Crossa, J., & Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian Linear Regression Package in R. *Plant Genome*, 3(2), 106-116, doi:10.3835/plantgenome2010.04.0005.
- Ramírez-Valverde, R., Hernández-Álvarez, O., Núñez-Domínguez, R., Ruíz-Flores, A., & García-Muñiz, J. G. (2007). Análisis univariado vs multivariado en la evaluación genética de variables de crecimiento en dos razas bovinas. *Agrociencia*, 41, 271-282.
- Ramírez-Valverde, R., Misztal, I., & Bertrand, J. K. (2001). Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. *Journal of Animal Science*, 79, 333-338.
- Ramírez-Valverde, R., Núñez-Domínguez, R., Ruíz-Flores, A., García-Muñiz, J. G., & Magaña-Valencia, F. (2008b). Comparación de definiciones de grupos contemporáneos en la evaluación genética de bovinos Suizo Europeo. *Técnica Pecuaria México*, 46(4), 359-370.
- Ramírez-Valverde, R., Núñez-Domínguez, R., Ruíz-Flores, A., García-Muñiz, J. G., & Domínguez-Viveros, J. (2008a). Estabilidad de las evaluaciones genéticas en poblaciones mexicanas de bovinos para carne. *Técnica Pecuaria México*, 46(1), 13-24.
- Saatchi, M., McClure, M. C., McKay, S. D., Rolf, M. M., & Kim, J. W. (2011). Accuracies of genomic breeding values in American Angus beef cattle using k-means clustering for cross-validation. *Genetics Selection Evolution*, 43, 40, doi:10.1186/1297-9686-43-40.
- Saatchi, M., Ward, J., & Garrick, D. J. (2013). Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *Journal of Animal Science*, 91, 1538-1551, doi:10.2527/jas.2012-5593.
- Saavedra-Jiménez, L. A., Ramírez-Valverde, R., Núñez-Domínguez, R., García-Muñiz, J. G., López-Villalobos, N., & Ruíz-Flores, A. (2013). Genotype by climate interaction in the genetic evaluation for growing traits of Braunvieh cattle in Mexico. *Tropical Animal Health and Production*, 45(7), 1489-1494, doi:10.1007/s11250-013-0387-2.
- Sargolzaei, M., Chesnais, J. P., & Schekel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15(1), 1-12, doi:10.1186/1471-2164-15-478.

- Schaeffer, L.R. (1984) Sire and cow evaluation under multiple trait models. *Journal of Dairy Science*, 67, 1567-1580.
- Schaeffer, L. R. (1994). Multiple-country comparison of dairy sires. *Journal of Dairy Science*, 77, 2671-2678.
- Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123, 218-223.
- Schefers, J., & Weigel, K. (2012). Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers*, 2(1), 4-9, doi:10.2527/af.2011-0032.
- Simianer, H. (2016). Genomic and other revolutions-why some technologies are quickly adopted and others are not. *Animal Frontiers*, 6(1), 53-58, doi:10.2527/af.2016-0008.
- Solarte-Portilla, C., & Rosero-Galindo, C. (2008). Aplicaciones actuales y potenciales de genética. Introducción Animal. *Revista de la facultad de ciencias económicas y administrativas. Universidad de Nariño*, 9(2), 109-129.
- Thallman, R. M., Hanford, K. J., Quaas, R. L., Kachman, S. D., Tempelman, R. J., Fernando, R. L., . . . Pollack, E. J. (2009). *Estimation of the proportion of genetic variation accounted for by DNA tests. Proceedings of the Beef Improvement Federation 41st Annual Research Symposium*. Sacramento, California, USA.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Statistical Society*, 58(1), 267-288.
- VanRaden, P. M., & Sullivan, P. G. (2010). International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution*, 42, 7, doi:10.1186/1297-9686-42-7.
- VanRaden, P. M., Van Tasell, G. R., Wiggans, T. S., Sonstergard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24, doi:10.3168/jds.2008-1514.
- Vazquez, A., Rosa, G., Weigel, K., de los Campos, G., Gianola, D., & Allison, D. (2010). Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. *Journal of Dairy Science*, 93(12), 5942-5949, doi:10.3168/jds.2010-3335.
- Wiggans, G. R., VanRaden, P. M., & Cooper, T. A. (2011). The genomic evaluation system in the United States: past, present, future. *Journal of Dairy Science*, 94, 3202-3211.
- Wiggans, G., Cole, J., Hubbard, S., & Sonstegard, T. (2017). Genomic selection in dairy cattle: The USDA experience. *Annual Review of Animal Biosciences*, 5, 309:327, doi:10.1146/annurev-animal-021815-111422.

- Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., . . . Dekkers, J. C. (2011a). Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution*, 43, 23, doi:10.1186/1297-9686-43-23.
- Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., . . . Dekkers, J. C. (2011b). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution*, 43, 5.

3 PREDICCIÓN DE VALORES GENÉTICOS CON UN MODELO ANIMAL Y BASADOS EN MARCADORES SNP PARA CARACTERÍSTICAS DE CRECIMIENTO EN BOVINOS SUIZO EUROPEO

3.1 Resumen

El objetivo fue comparar las predicciones de valores genéticos (VG) de pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo europeo obtenidas con un modelo animal (PBLUP) utilizando información fenotípica y de pedigrí con las obtenidas con modelos basados en marcadores moleculares: GBLUP (PBLUP+información genómica), GBLUP+A (PBLUP+información genómica y de pedigrí) y RKHS (Reproducing Kernel Hilbert Space). Los criterios para comparar los modelos fueron: magnitud de los estimadores de parámetros genéticos, habilidad de predicción, jerarquización de los VGs predichos y estimadores de los coeficientes de regresión de los VGs predichos con un modelo sobre los obtenidos con otro. Los modelos genómicos rindieron mayores estimadores de los componentes de varianza genética y ambiental para las tres características evaluadas. Mediante validación cruzada se observó mejor habilidad predictiva para los tres modelos que incluyeron información genómica para las tres características estudiadas. La jerarquización y relación de los VGs predichos para PN y PA fue diferente entre BLUP y modelos genómicos, para PD cada uno de los modelos produjo predicciones y jerarquización de los VGs equivalentes a las de los demás. El cambio promedio de las predicciones de los VGs respecto a los fenotipos pre-ajustados fue mayor con PBLUP (1.322) que con GBLUP (1.193), GBLUP+A (1.185) y RKHS (1.205). Lo anterior indica que las predicciones con los modelos basados en marcadores moleculares, mejoraron las estimaciones de los parámetros genéticos, asociaciones de los valores observados y predichos, y redujeron el sesgo de los valores genómicos respecto a los valores fenotípicos pre-ajustados para algunas características de crecimiento comparadas con PBLUP. La implementación de la selección de los mejores animales con base en información genómica podría tener mayor impacto que la selección basada en información fenotípica y de pedigrí en Bovinos Suizo europeo.

Palabras clave: Evaluación genómica, Bovinos Suizo europeo, Modelos de evaluación genética

PREDICTION OF BREEDING VALUES WITH AN ANIMAL MODEL AND BASED ON SNP MARKERS FOR GROWTH TRAITS IN BRAUNVIEH CATTLE

3.2 Abstract

This study aimed to compare the breeding values (BV) for birth (BW), weaning (WW), and yearling (YW) weights of Braunvieh cattle, predicted with four models. Models compared were a PBLUP (Best Linear Unbiased Predictor), Animal Model using phenotypic and pedigree information, and three models using molecular marker data: GBLUP (BLUP+genomic information), GBLUP+A (BLUP+genomic and pedigree information), and RKHS (Reproducing Kernel Hilbert Space). The models were compared using the criteria: magnitude of genetic parameter estimates, prediction ability, ranking of predicted BVs, and regression estimates of BVs predicted with a model on the BVs predicted with another model. Genomic models yielded higher genetic and environmental variance components for the three evaluated traits than PBLUP. Cross-validation produced better predictive ability for the genomic models for the three traits evaluated. The ranking and relationships of the predicted BVs for BW and YW were different between PBLUP and genomic models. However, for WW each of the models produced predictions and ranking of BVs equivalent to the rest of the models. The average change of the predictions of the BV on pre-adjusted phenotypes was higher in PBLUP (1.322) than in GBLUP (1.193), GBLUP+A (1.185), and RKHS (1.205). This suggests that the predictions of the models based on molecular markers improved estimates of genetic parameters, associations of the pre-adjusted phenotypes and predicted BVs, and reduced the bias of the genomic values with respect to the pre-adjusted phenotypes for some traits compared with PBLUP. The selection of animals based on genomic information could have a greater impact than that based on phenotypic and pedigree information in Braunvieh cattle.

Key words: Genomic evaluation, Braunvieh cattle, Models for genetic evaluation

3.3 Introducción

Los bovinos productores de carne de la raza Suizo europeo se utilizan en México como raza pura o en sistemas de cruzamiento para producción de carne, se adaptan bien a regiones templadas y tropicales. En México, a partir de 2003 se realizan evaluaciones genéticas para ésta raza, basadas en fenotipos y pedigrí. Sin embargo, con el desarrollo de nuevas tecnologías, como la Selección Genómica (SG), es posible la predicción de valores genéticos para estos animales en etapas tempranas de su vida (Meuwissen et al., 2001).

La SG, permite disminuir los costos de los programas de selección de animales y el intervalo entre generaciones, e incrementar la exactitud de los valores genéticos predichos, además, de complementar la evaluación genética tradicional basada en pedigrí y fenotipos (Hayes et al., 2009; VanRaden et al., 2009; de los Campos et al., 2013). La predicción de los valores genéticos puede realizarse con una exactitud relativamente alta, por lo que se requiere que un número grande de SNPs (Nucleótidos de Polimorfismo Simple) cubran el genoma completo. El número de marcadores moleculares (p) suele ser mayor que el número de observaciones (n). La predicción de los valores genéticos por modelos convencionales de evaluación genética, considerando toda la información de marcadores, no es posible, es necesario un enfoque diferente para usar de manera eficiente toda la información genómica. Esto puede realizarse mediante alternativas como GBLUP, Regresión Ridge-BLUP (RR-BLUP), RKHS, Lasso Bayesiano, BayesA, BayesB y BayesC, entre otros (de los Campos et al., 2013).

Meuwissen et al. (2001) compararon la metodología BLUP con análisis Bayesianos en la predicción de valores genéticos en individuos genotipados y sin registros fenotípicos de una población simulada. Estos autores encontraron que los métodos bayesianos incrementaron la exactitud de las predicciones hasta 0.85, comparada con BLUP donde la exactitud fue 0.73. El método BayesA resultó con 9% más de exactitud que BLUP y la exactitud de BayesB excedió 16% a BLUP.

Éstas metodologías pueden diferir en magnitud y exactitud de los valores genéticos predichos. Cada una de ellas se adecuará mejor dependiendo de las condiciones específicas de la población a evaluar, i. e. tamaño y estructura de la población a estudiar (población de entrenamiento y de evaluación), densidad del panel de marcadores SNP y confiabilidad de la información fenotípica, entre otros factores. El objetivo del estudio fue comparar las predicciones de valores genéticos para características de crecimiento obtenidas con el modelo animal y con modelos basados en marcadores SNP en bovinos Suizo europeo.

3.4 Materiales y métodos

3.4.1 Origen y edición de la información

3.4.1.1 Fenotípica

Los registros genealógicos y productivos utilizados en el estudio pertenecen a la Asociación Mexicana de Criadores de Ganado Suizo de Registro. El pedigrí lo conformaron 184,788 animales nacidos entre 1901 y 2016. Los registros de comportamiento productivo correspondieron a animales nacidos entre 1985 y 2016 de 240 ranchos ubicados en varios estados de la República Mexicana.

Las variables que se analizaron fueron los pesos al nacimiento (PN), al destete (PD) y al año (PA). En la edición de todas las características, animales sin registros de pesos, fechas al nacimiento, al destete y al año, hato, manejo y edad de la madre fueron descartados. Las observaciones fuera del intervalo $\bar{X} \pm 3\sigma$ también se eliminaron.

Las épocas de nacimiento se definieron considerando la cantidad y distribución de precipitación y el promedio de temperatura mensual a través de los años de la estación meteorológica más cercana a cada rancho (SMN, 2017). Se definieron cuatro épocas de nacimiento: primavera, verano, otoño e invierno.

La condición de manejo para PD se definió de la siguiente manera: 1) alimentación de la cría con leche de la madre; 2) alimentación de la cría con leche

de la madre más alimento balanceado; y 3) alimentación de la cría con leche de la madre y de una nodriza más alimento balanceado. La condición de manejo para PA se definió considerando los siguientes criterios: 1) pastoreo, sales minerales a libre acceso y ensilado o melaza en la época seca; 2) pastoreo y sales minerales, más una dieta balanceada; y 3) estabulados.

Los grupos contemporáneos (GC) se formaron combinando los efectos de rancho, año y época de nacimiento con un mínimo de dos observaciones por grupo. Los GC con más de tres observaciones que no presentaron variabilidad dentro de ellos en los datos productivos también fueron descartados. La conexión entre GC se determinó con el programa AMC (Roso & Schenkel, 2006) y se descartaron animales en GC no conectados. En el Cuadro 3 se presenta el número final de observaciones, grupos contemporáneos y estadísticos descriptivos de peso al nacimiento, al destete y al año, después de la edición de los datos y que fueron utilizados para el pre-ajuste.

Cuadro 3. Estadísticos descriptivos, número de observaciones (n) y número de grupos contemporáneos (GC) utilizados en el pre-ajuste de los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo europeo.

Característica	Mínimo	Máximo	$\bar{X} \pm \sigma$	CV ^z (%)	N	GC
PN, kg	24.0	52.0	38.09±4.86	12.77	28900	3047
PD, kg	103.0	337.0	218.37±37.28	17.07	24360	2729
PA, kg	135.0	507.0	320.17±60.14	18.78	17393	1996

^z Coeficiente de variación

3.4.1.2 Genómica

Para la utilización de las metodologías basadas en marcadores moleculares, se obtuvieron muestras de folículos pilosos de 150 animales Suizo europeo (Cuadro 4). Las muestras se genotiparon en la empresa Geneseek Neogen Co., Lincoln, NE, USA. El procedimiento de genotipado se realizó mediante el chip GGP Bovine LD 30K con un promedio de espacio entre marcadores de 89k pares de bases.

Cuadro 4. Número y procedencia de los animales utilizados para la obtención de muestras de ADN.

Rancho	Machos	Hembras	Total	Municipio y estado
1	0	32	32	Casitas, Veracruz
2	11	35	46	Etzalan, Jalisco
3	18	32	50	Lagos de Moreno, Jalisco
4	5	2	7	Villa de Álvarez, Colima
5	5	10	15	Papantla, Veracruz

El chip para el genotipado tuvo 30,106 SNPs. Los genotipos se obtuvieron en el formato de alelos Illumina A/B y se codificaron como 0, 1 o 2, que representa el número de alelos B. Debido a que no se dispuso de información para todos los genotipos, se hizo una imputación con el paquete BGLR del programa R (Pérez-Rodríguez & de los Campos, 2014). Se encontraron 214 SNPs con valores faltantes, 1,587 monomórficos y 2,098 SNPs tuvieron una frecuencia de alelo menor que 0.05, los cuales no fueron incluidos en el análisis. De los 30,106 SNPs iniciales, 26,207 se consideraron para el análisis genómico después del control de calidad.

3.4.2 Análisis de la información

Para la evaluación tradicional y genómica se pre-ajustaron los registros productivos por efectos no genéticos. El modelo para pre-ajustar sólo incluyó el efecto aleatorio de GC. Posteriormente, se obtuvieron los fenotipos pre-ajustados de los animales genotipados y se utilizaron como variable de respuesta en los análisis, donde se incluyó el resto de los efectos como fijos o aleatorios con los que se obtuvo mejor ajuste para cada variable (Cuadro 5).

El pre-ajuste se realizó debido a que el comportamiento productivo de los animales está afectado por efectos no genéticos como sexo y edad del animal, año y época de nacimiento, etc., por lo cual es importante considerarlos para reducir la variabilidad entre animales debida a causas no genéticas. Además, de ésta manera los requerimientos computacionales son menores. Con el pre-ajuste de la información fenotípica considerando estos efectos, ésta queda sin la influencia de dichos efectos, y por tanto, las estimaciones de los efectos

genéticos quedan libres de ellos. Inicialmente, cuando se pre-ajustó la información fenotípica simultáneamente por diversos efectos, y posteriormente, al modelar el efecto de marcadores, las correlaciones entre valores genéticos predichos y fenotipos pre-ajustados fueron cercanas a la unidad, lo que indicó sobreajuste. Sin embargo, las inconsistencias persistieron (Figura 6). Para reducir el sobreajuste se pre-ajustó sólo por GC. Los análisis finales consideraron el pre-ajuste de los fenotipos sólo por GC como efecto aleatorio, el resto de los efectos se consideraron en la modelación genómica como se especifican en el Cuadro 5.

Cuadro 5. Efectos considerados en los modelos de pre-ajuste.

Característica	Fijo		Aleatorio	Covariable (lineal y cuadrática)		
	Sexo	Condición de Manejo	GC	Edad de la madre	Edad al destete	Edad al año
PN	✓		✓	✓		
PD	✓	✓	✓	✓	✓	
PA	✓	✓	✓			✓

GC= Grupo contemporáneo; PN=Peso al nacimiento; PD= Peso al destete y PA= Peso al año.

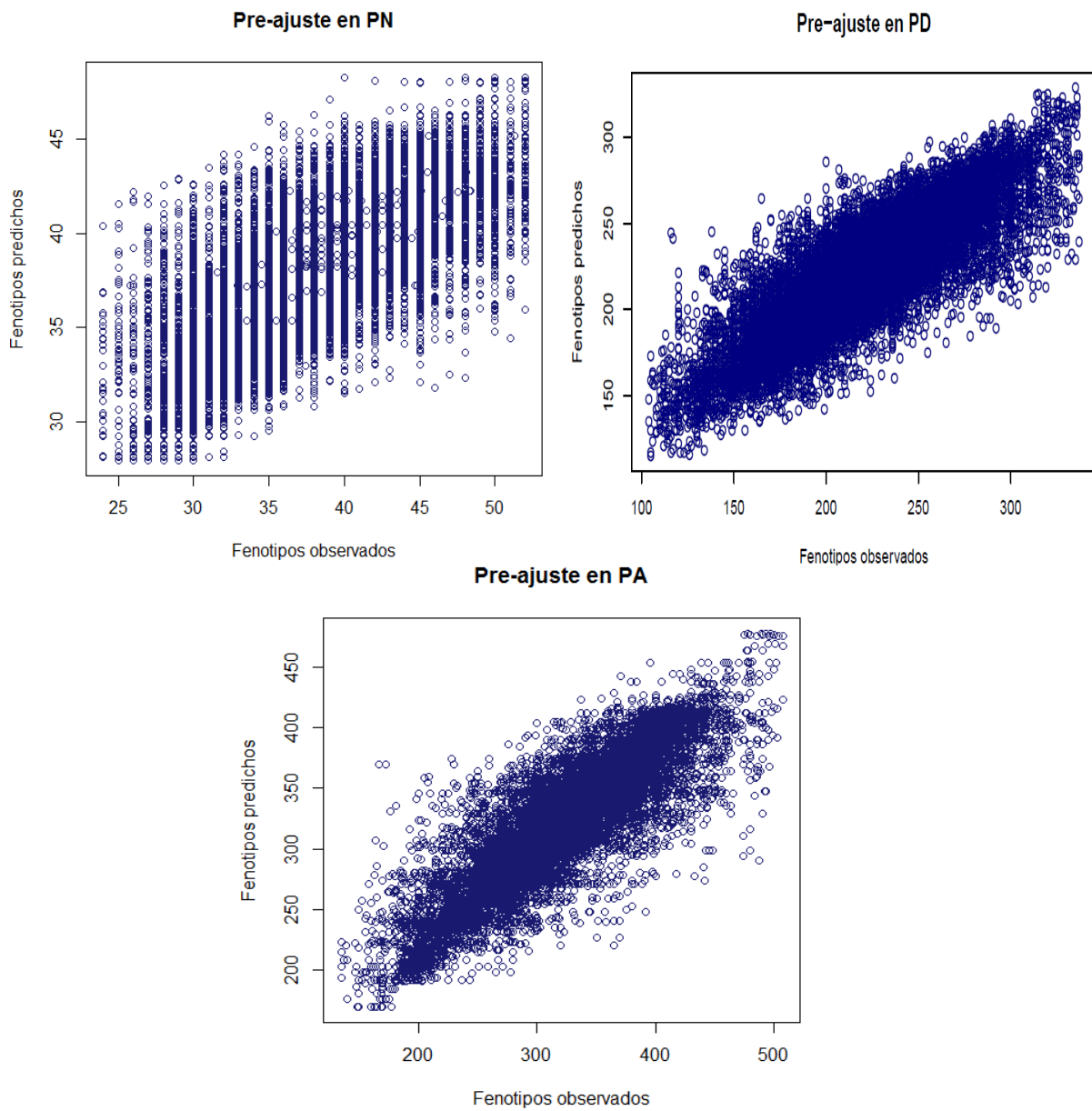


Figura 6. Distribución de datos productivos de peso al nacimiento, al destete y al año después de pre-ajustarlos por efectos no genéticos.

3.4.2.1 Metodología cuantitativa tradicional (Modelo Animal)

Después del pre-ajuste, se obtuvieron los fenotipos pre-ajustados de los animales genotipados y se utilizaron como variable de respuesta en los análisis. El número final de observaciones utilizadas para la predicción de valores genéticos de cada una de las variables analizadas se presenta en el Cuadro 6.

Cuadro 6. Estadísticos descriptivos y número de observaciones (n) utilizados en los análisis genéticos para los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo europeo.

Característica	Mínimo	Máximo	$\bar{X} \pm \sigma$	CV ^z (%)	n
PN (kg)	34.00	44.10	38.53±3.11	8.09	120
PD (kg)	189.00	251.73	216.14±19.52	9.03	105
PA(kg)	244.12	383.33	320.91±28.44	8.86	95

^z Coeficiente de variación.

Los análisis estadísticos para predecir los VGs y estimar los parámetros genéticos, se realizaron utilizando el modelo animal con el paquete BGLR (Pérez-Rodríguez & de los Campos, 2016) del programa R. Las tres variables de respuesta se analizaron con modelos univariados. Lo anterior, fue por limitaciones del programa para realizar análisis bivariados. El modelo general fue:

BLUP tradicional con información de pedigrí (PBLUP)

$$y = \mu + Xb + Zg + Mm + \varepsilon \dots (1)$$

donde, y es el vector de registros fenotípicos pre-ajustados para cada variable de crecimiento, μ es la media, b es el vector de efectos fijos, g es el vector de efectos genéticos aditivos directos, m es el vector de efectos genéticos directos maternos (únicamente para PD), y ε es el vector de efectos residuales. X , Z y M son matrices de incidencia que relacionan las observaciones con los vectores correspondientes. La estructura de (co)varianzas para los efectos aleatorios fue:

$$V \begin{bmatrix} u \\ m \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & A\sigma_{am} & 0 \\ A\sigma_{am} & A\sigma_m^2 & 0 \\ 0 & 0 & I\sigma_\varepsilon^2 \end{bmatrix}$$

donde:

σ_a^2 es la varianza genética aditiva directa, σ_m^2 es la varianza genética aditiva materna, σ_{am} es la covarianza entre efectos genéticos aditivos directo y materno

y **A** es la matriz de relaciones genéticas aditivas entre los animales derivada del pedigrí.

3.4.2.2 Metodologías basadas en marcadores SNP

Después del pre-ajuste, se obtuvieron los fenotipos pre-ajustados de los animales genotipados y se utilizaron como variable de respuesta en los análisis. El Cuadro 6 presenta estadísticos descriptivos de las variables para las cuales se predijeron los valores genómicos.

Las predicciones de valores y parámetros genómicos se realizaron utilizando el paquete BGLR (Pérez & de los Campos, 2014) del programa R. Se usaron tres modelos para estimar los efectos de marcadores: GBLUP, GBLUP+A y RKHS. El modelo utilizado incluyó efectos fijos (Cuadro 5), efectos genéticos determinados por marcadores y pedigrí y el efecto materno, este último únicamente para PD.

BLUP con uso de marcadores (GBLUP)

$$y = \mu + Xb + Zg + Mm + \varepsilon \dots (2)$$

donde y_i es el vector de registros fenotípicos pre-ajustados de las variables de crecimiento analizadas; μ es la media, b es el vector de efectos fijos, g es el vector de efectos genéticos aditivos directos, determinado por información de marcadores y m es el vector de efecto genético materno (únicamente para PD), X, Z y M son matrices de incidencia que relacionan las observaciones con los vectores correspondientes y ε es el residual.

La suposición fue que $Zg = u_1$, u_1 es el efecto genético para el i -ésimo animal, determinado por información de marcadores, $u_1 \sim N(0, G\sigma_u^2)$, y $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$.

GBLUP más pedigrí (GBLUP+A)

$$y = \mu + Xb + Zg + Wp + Mm + \varepsilon \dots (3)$$

Los términos similares al modelo 2, se definen de igual manera, y ahora se asume que $Wp = u_2$, u_2 es el efecto genético para el i -ésimo animal, determinado por información de pedigrí, $u_2 \sim N(0, A\sigma_u^2)$, donde A es la matriz de relaciones genéticas aditivas derivada del pedigrí.

Reproducing Kernel Hilbert Space (RKHS)

$$y = \mu + Xb + K_h\alpha + Mm + \varepsilon \dots (4)$$

Los términos similares a los modelos anteriores, se definen de igual manera, y ahora: K_h es una matriz de incidencias, que depende de la elección del kernel y del parámetro h , que de forma semejante a una matriz de correlaciones mide la similitud genómica entre los genotipos. Se supone que $\alpha \sim N(0, K_h \sigma_\alpha^2)$, donde K se calcula de la siguiente manera $K_h = \{x_i, x_{i'}\} = \exp \left\{ -h \frac{\sum_{k=1}^p (x_{ik} - x_{i'k})^2}{p} \right\}$.

3.4.3 Criterios de comparación entre metodologías

Los criterios para comparar los resultados de los análisis fueron:

Magnitud de los estimadores de los parámetros genéticos: el primer criterio utilizado para la comparación de metodologías fue la magnitud de los estimadores de heredabilidad.

Habilidad de predicción: la comparación de los cuatro modelos se realizó mediante validación cruzada con 50 repeticiones. Para ello, el conjunto de datos se dividió aleatoriamente en grupos, se omitió 20% de los valores fenotípicos de las variables evaluadas; luego se obtuvieron los predictores de los VGs para los registros eliminados y se obtuvieron los coeficientes de correlación de los fenotipos eliminados con sus VGs predichos. Adicionalmente, se obtuvieron los cuadrados medios del error (CME), calculados como el promedio de las diferencias al cuadrado entre el valor real y el predictor obtenido con cada modelo.

Jerarquización de los VGs predichos de los animales: para determinar el grado de similitud en la jerarquización de los VGs predichos con los diferentes modelos, se estimaron: 1) correlaciones Pearson y Spearman entre los VGs predichos con un modelo con los obtenidos con cada uno de los restantes para PN, PD y PA; y 2) porcentaje de coincidencias de los 20 animales superiores

ordenados con respecto a sus VG predichos con cada uno de los cuatro modelos probados para las características evaluadas.

Coefficiente de regresión lineal simple: para determinar los cambios en la magnitud de los VGs predichos, a) con respecto a los fenotipos pre-ajustados, y b) entre cada uno de los modelos, se obtuvieron los coeficientes de regresión lineal simple. Este coeficiente permite evaluar el sesgo de las estimaciones entre cada uno de los modelos. La hipótesis que se probó fue que el estimador del coeficiente de regresión no fue diferente de cero.

3.5 Resultados y discusión

3.5.1 Magnitud de los estimadores de los parámetros genéticos

Los componentes de varianza y heredabilidades directa y materna estimados con cada modelo se muestran en el Cuadro 7. Los componentes de varianza genética aditiva directa para PN, PD y PA obtenidos con los tres modelos genómicos fueron mayores que el obtenido con PBLUP. Por el contrario, los mismos modelos rindieron menores estimadores de los componentes de varianza genética materna y fenotípica en comparación con el PBLUP.

La heredabilidad directa fue mayor para GBLUP, GBLUP+A y RKHS comparada con la obtenida con PBLUP para las tres características estudiadas. Con respecto a la heredabilidad materna de PD, con el modelo PBLUP se obtuvo el mayor estimador. De acuerdo con Meuwissen et al. (2001), el incremento de los estimadores de heredabilidad directa obtenidos con los modelos genómicos con respecto al modelo tradicional, podrían atribuirse a una mayor captura de variabilidad fenotípica para las características de crecimiento evaluadas. Sin embargo, estos resultados no son consistentes en la literatura. Los modelos genómicos, por su cantidad grande de variables independientes, pueden ser sobre-ajustados, o porque el efecto para un SNP puede estar fuertemente influido por la presencia o ausencia de un SNP vecino en desequilibrio de ligamiento con el SNP de interés (Weigel et al., 2009). Es por esto que se ha propuesto incluir en el panel de marcadores todas las variantes causales que influyen las

características en evaluación para inferir de mejor manera sobre sus parámetros genéticos (Speed, Hemani, Johnson & Balding, 2012).

Cuadro 7. Estimadores de parámetros genéticos para características de crecimiento en bovinos Suizo europeo.

Estimador	Modelo	Característica		
		PN	PD	PA
σ_a^2	PBLUP	2.50	31.52	136.75
	GBLUP	4.21	47.28	225.61
	GBLUP+A	4.61	61.78	275.19
	RKHS	5.71	35.13	153.98
σ_m^2	PBLUP	-	30.07	-
	GBLUP	-	25.94	-
	GBLUP+A	-	23.78	-
	RKHS	-	23.40	-
σ_p^2	PBLUP	8.11	174.11	561.52
	GBLUP	6.31	168.86	503.93
	GBLUP-A	6.63	173.35	521.82
	RKHS	7.87	151.03	433.03
h_a^2	PBLUP	0.31	0.18	0.24
	GBLUP	0.67	0.28	0.45
	GBLUP-A	0.70	0.36	0.53
	RKHS	0.73	0.23	0.36
h_m^2	PBLUP	-	0.17	-
	GBLUP	-	0.15	-
	GBLUP-A	-	0.14	-
	RKHS	-	0.15	-

σ_a^2 = Varianza genética directa; σ_m^2 = Varianza genética materna; σ_p^2 = varianza fenotípica total; h_a^2 = heredabilidad directa; h_m^2 = heredabilidad materna; h_p^2 = heredabilidad materna; PBLUP= BLUP tradicional usando información de pedigrí; GBLUP: BLUP con información de marcadores; GBLUP+A: BLUP con información de marcadores y pedigrí; RKHS: *Reproducing Kernel Hilbert Space*; PN= Peso al nacimiento; PD = Peso al destete y PA= Peso al año.

El incremento en la estimación de heredabilidad directa obtenida con modelos genómicos se ha evaluado en otras razas de bovinos productores de carne. Ryoo y Lee (2014) estimaron heredabilidades de 0.58 a 0.76 para pesos corporales de ganado bovino coreano a diferentes edades, utilizando información de 39,112 marcadores SNP. Por el contrario, Ríos-Utrera (2008) estimó menores estimadores para PN tanto de heredabilidad directa como materna, obtenidos con

metodologías tradicionales, 0.34, 0.32, 0.38, 0.43, 0.29, 0.34 y 0.36; y 0.17, 0.10, 0.14, 0.20, 0.08, 0.09 y 0.11 para Angus, Brahman, Charolais, Hereford, Limousin, Nelore y Simmental, respectivamente.

3.5.2 Habilidad de predicción

Un estimador mayor para el coeficiente de correlación Pearson entre los VGs de los animales eliminados en la población de validación y sus fenotipos, implica una mayor habilidad predictiva del modelo. Esto quiere decir que es el mejor modelo para predecir VGs.

El modelo PBLUP presentó menor habilidad predictiva comparada con los modelos genómicos para cada una de las características evaluadas cuando se utilizó validación cruzada (Cuadro 8). Los estimadores de correlación Pearson entre los VGs de los animales eliminados y sus fenotipos fueron 0.452, 0.759 y 0.543 para PN, PD y PA, respectivamente. La habilidad predictiva de los modelos genómicos aumentó en promedio 65.0, 5.5 y 55.0% para PN, PD y PA, respectivamente, comparados con PBLUP. Los cambios en habilidad predictiva dentro de modelos genómicos para cada característica, fueron relativamente pequeños.

Cuadro 8. Promedios de los estimadores de correlación Pearson y cuadrados medios del error (CME) obtenidos para diferentes modelos utilizando validación cruzada de 50 repeticiones para PN, PD y PA de Bovinos Suizo europeo.

Modelos	Correlación			CME		
	PN	PD	PA	PN	PD	PA
PBLUP	0.452	0.759	0.543	11.763	166.924	589.928
GBLUP	0.774	0.824	0.672	4.514	124.230	451.879
GBLUP+A	0.765	0.795	0.688	4.471	150.672	433.997
RKHS	0.704	0.782	0.621	5.200	155.479	513.943

PBLUP = BLUP tradicional usando información de pedigrí; GBLUP = BLUP con información de marcadores; GBLUP+A = BLUP con información de marcadores y pedigrí; RKHS = *Reproducing Kernel Hilbert Space*; PN = Peso al nacimiento; PD = Peso al destete y PA = Peso al año.

Los valores altos de los estimadores de correlación obtenidos, podrían indicar una mayor contribución de las relaciones genéticas aditivas entre las poblaciones

de entrenamiento y validación capturadas por el número grande de marcadores SNPs del panel utilizado. Estos cambios son consistentes con los observados en otros estudios (Meuwissen et al., 2001; Habier et al., 2007, VanRaden & Sullivan 2010) y reflejan la diferencia por el tipo de información utilizada. Cuando se utiliza la matriz A de relaciones genéticas aditivas derivada del pedigrí (PBLUP), los estimadores de correlación son menores, debido a que A considera la fracción esperada de genes idénticos por descendencia, mientras que los modelos donde que utilizan la matriz G de marcadores, consideran la fracción real de ADN compartido por los individuos (Goddard, Hayes & Meuwissen, 2011, de los Campos et al., 2009).

La ventaja de los modelos genómicos fue más notoria para para PN y PA, cuando se incluyó información de marcadores y pedigrí, donde se obtuvieron las mayores correlaciones y los menores CME. Mientras que para PD el mejor modelo fue GBLUP, Cuadro 8. La inclusión de información de marcadores y pedigrí simultáneamente en el modelo es importante, de esta manera se incrementa el tamaño de la población de entrenamiento sin la necesidad de genotipar más animales. Lo anterior, implica que animales con registros productivos y sin información genómica disponible pueden evaluarse conjuntamente con individuos con información genómica mediante el uso de información proveniente del pedigrí (Misztal, Legarra & Aguilar, 2009; Christensen & Lund, 2010).

En varios estudios se ha documentado la mejor habilidad predictiva de los modelos genómicos. Gonzáles-Recio et al. (2008) evaluaron la capacidad predictiva de cuatro modelos mediante validación cruzada para mortalidad de aves entre 14 y 42 días de edad, los modelos evaluados fueron PBLUP, RKHS, Regresión Bayesiana y Regresión Kernel. Esto autores encontraron que RKHS tuvo mejor ajuste, menor CME y 25% de incremento en la precisión en comparación con los demás modelos. Los autores concluyeron que RKHS, usando pocos SNPs tiene mejor capacidad predictiva que Regresión Bayesiana al detectar una mejor asociación de los SNP con la característica evaluada.

En otro estudio, donde se evaluó la habilidad predictiva de siete modelos incluyendo PBLUP, RKHS y GBLUP, los estimadores de correlación entre fenotipos observados y predichos para tamaño de camada al nacimiento de tres líneas de cerdos, fueron mayores en RKHS y GBLUP comparados con PBLUP (Tusell et al., 2011). Estos resultados muestran que el rendimiento superior de los modelos basados en marcadores moleculares es mayor para características de baja heredabilidad, como tamaño de camada al nacimiento. En otro estudio, Calus et al. (2014) compararon un modelo convencional con información de pedigrí vs modelos de predicción genómica (GBLUP, RR-BLUP y BayesC) en la exactitud de las predicciones de VGs para número de huevos en el primer periodo de postura en gallinas ponedoras. Los modelos de predicción genómica tuvieron una precisión de 0.13 a 0.16 más alta que BLUP basado en pedigrí.

Los cambios en capacidad predictiva de los modelos genómicos fueron pequeños en el presente estudio, por lo que es importante la elección de éstos, pues se ha demostrado, con información simulada, que esto afecta la capacidad predictiva, debiéndose principalmente a la arquitectura genética de la característica, esto es, del número de QTL's, su distribución y sus efectos en las características de interés, así como al tamaño del panel de marcadores y tamaño de muestra (Habier et al., 2007; Hayes et al., 2009; Pszczola, Strabel, Mulder & Calus, 2012). La precisión es mayor para modelos de selección de variables como BayesB y BayesC que se ajustan mejor a la arquitectura genética de la característica frente a los modelos GBLUP y GBLUP+A (Bastiaansen et al., 2010; Pszczola et al., 2012). Además, con modelos que utilizan paneles de marcadores de baja densidad puede ser poco probable que los marcadores estén estrechamente ligados a QTLs. Lo anterior induce una distribución menos extrema de los efectos que los obtenidos con paneles de mayor densidad (de los Campos et al., 2013). Por otro lado, para cualquier densidad de marcadores, la habilidad predictiva de los modelos genómicos aumenta a medida que se incrementa el desequilibrio de ligamiento entre marcador y QTL. El incremento en el tamaño de muestra se asocia con un aumento en el alcance de las relaciones genéticas entre individuos de las poblaciones de entrenamiento y validación. Lo anterior indica que entre

mayor sea el panel de marcadores, el tamaño de muestra también deberá ser más grande, puesto que con un tamaño de muestra pequeño, deberá tenerse cuidado en la interpretación de los efectos de marcadores porque estos podrían estar influidos por la elección de información a priori (de los Campos et al., 2013).

Con información real, la arquitectura genética es menos conocida, lo que implica que hay pocas características para las que uno o pocos genes explican una proporción grande de la varianza genética, por lo que los modelos no se ven limitados para expresar su potencial predictivo (Daetwyler, Pong-Wong, Villanueva & Woolliams, 2010). Sin embargo, para PN, PD y PA, existen varios genes relacionados con éstas características (Buzanskas et al., 2014; Kneeland et al., 2004; Snelling et al., 2014), por lo que es necesario considerar todas las variantes causales y evaluar otros modelos para obtener predicciones más confiables.

3.5.3 Jerarquización de los VGs predichos de los animales

En el Cuadro 9 se muestran los estimadores de correlaciones Pearson y Spearman entre los VGs predichos para los cuatro modelos evaluados. Las correlaciones fueron altas entre todos los modelos diferentes de PBLUP, no así entre PBLUP y los genómicos, las correlaciones Pearson oscilaron de 0.883 a 0.889 para PN, de 0.990 a 0.992 para PD y de 0.930 a 0.947 para PA. Mientras que las correlaciones Spearman variaron de 0.931 a 0.939, 0.974 a 0.981 y 0.938 a 0.953 para PN, PD y PA, respectivamente. No hubo una diferencia notoria entre los modelos, con los genómicos se obtuvieron los mayores estimadores de correlación Pearson y Spearman.

Cuadro 9. Estimadores de los coeficientes de correlación Pearson (arriba de la diagonal) y Spearman (debajo de la diagonal), entre paréntesis nivel de significancia para la hipótesis del estimador igual a cero, entre valores genéticos predichos con cuatro modelos de evaluación genética en bovinos Suizo europeo.

Modelos	PBLUP	GBLUP	GBLUP+A	RKHS
PN (<0.0001)				
PBLUP		0.889	0.882	0.883
GBLUP	0.931		0.999	0.999
GBLUP+A	0.939	0.996		0.999
RKHS	0.935	0.996	0.997	
PD (<0.0001)				
PBLUP		0.991	0.990	0.992
GBLUP	0.974		0.995	0.998
GBLUP+A	0.981	0.983		0.998
RKHS	0.979	0.992	0.991	
PA (<0.0001)				
PBLUP		0.936	0.930	0.947
GBLUP	0.939		0.995	0.996
GBLUP+A	0.938	0.989		0.996
RKHS	0.953	0.992	0.987	

PBLUP= BLUP tradicional usando información de pedigrí; GBLUP: BLUP con información de marcadores; GBLUP+A: BLUP con información de marcadores y pedigrí; RKHS: *Reproducing Kernel Hilbert Space*; PN= Peso al nacimiento; PD= Peso al destete y PA= Peso al año.

Los porcentajes de coincidencia entre los VGs predichos con cada uno de los modelos evaluados para los 20 animales superiores, se muestran en el Cuadro 10. Para PN, el porcentaje promedio de coincidencias entre los VGs predichos con PBLUP y los modelos genómicos fue menor (55%) que para PD (85%) y PA (71.6%). El porcentaje de coincidencias entre modelos genómicos varió de 85 a 100%, considerando todas las características. Lo anterior indica que la jerarquización y relación de los de los VGs para PN y PA de los animales evaluados, es sensible al método utilizado para predecirlos. Sin embargo, para PD, todos los modelos produjeron predicciones y jerarquizaciones similares de los VGs de los animales.

Cuadro 10. Porcentajes de coincidencia de los 20 animales superiores ordenados respecto a los valores genéticos predichos con cuatro modelos de evaluación genética para pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo europeo.

Métodos	PBLUP	GBLUP	GBLUP+A	RKHS
PN				
PBLUP	100	55	55	55
GBLUP		100	100	100
GBLUP+A			100	100
RKHS				100
PD				
PBLUP	100	80	90	85
GBLUP		100	85	95
GBLUP+A			100	80
RKHS				100
PA				
PBLUP	100	70	70	75
GBLUP		100	95	95
GBLUP+A			100	85
RKHS				100

PBLUP = BLUP tradicional usando información de pedigrí; GBLUP = BLUP con información de marcadores; GBLUP+A= BLUP con información de marcadores y pedigrí; RKHS = *Reproducing Kernel Hilbert Space*.

De los Campos et al. (2009) analizaron información del índice de masa corporal de una población de ratones (n=1884) genotipados con 10,946 marcadores y evaluaron tres modelos para predecir VGs: 1) basado en pedigrí (P), 2) basado en marcadores (M), y 3) basado en marcadores y pedigrí (MP). Estos autores determinaron los coeficientes de correlación Spearman entre los VGs predichos, y encontraron estimadores que variaron de 0.715 a 0.986 para los tres modelos evaluados. La correlación entre los VGs predichos con M y MP fue mayor que la correspondiente a P y MP, lo que sugiere que la inclusión de marcadores permite la captura de fracciones de varianza genética aditiva que se perderían si los pedigríes fueran la única fuente de información genética utilizada, y que los modelos que incluyen información de marcadores y pedigrí tienen mejor habilidad predictiva que los basados en pedigrí o marcadores. En un estudio realizado en tres líneas de gallinas, Calus et al. (2014) compararon los VGs predichos con los modelos PBLUP, RR-BLUP y GBLUP para número de huevos en el primer ciclo de postura. Estos autores encontraron que los estimadores de correlación

Pearson promedio entre los VGs predichos y los fenotipos observados de las tres líneas fueron 0.59, 0.60 y 0.96 para PBLUP y GBLUP, PBLUP y RR-BLUP, y GBLUP y RR-BLUP, respectivamente. Los modelos GBLUP y RRBLUP son matemáticamente equivalentes, por eso el coeficiente de correlación obtenido es cercano a la unidad. Por tanto, estos dos modelos realizan predicciones de VGs equivalentes. Además, estos autores estimaron los coeficientes de regresión de los fenotipos sobre los VGs para cada uno de los modelos, los promedios para las tres líneas fueron 1.20, 1.06 y 1.24 para PBLUP, GBLUP y RR-BLUP, respectivamente.

3.5.4 Coeficiente de regresión lineal simple

Los coeficientes de regresión de los valores fenotípicos pre-ajustados para las variables analizadas sobre los VGs predichos con los modelos estudiados se muestran en el Cuadro 11. Estimadores de estos coeficientes igual a la unidad, indican que los modelos predicen los VGs de manera similar. Si bien no hubo mucha diferencia entre los estimadores obtenidos con los modelos comparados, considerando todas las variables estudiadas, el correspondiente a PN estimado con PBLUP fue el que mostró un cambio mayor comparado con los demás.

Wolc et al. (2011) estimaron coeficientes de regresión de peso corporal y altura de albúmina sobre sus VGs predichos en gallinas de postura de 1.12 y 1.29 para PBLUP, 1.10 y 1.06 para GBLUP y 1.06 y 1.013 para BayesCπ, respectivamente. Estos resultados indican que los VGs predichos GBLUP y BayesCπ, son más parecidos entre sí, que los predichos con PBLUP.

Cuadro 11. Estimadores de los coeficientes de regresión ($\pm$ error estándar), de los valores fenotípicos pre-ajustados sobre valores genéticos predichos para pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo europeo.

Modelos	PBLUP	GBLUP	GBLUP+A	RKHS
PN	1.551 $\pm$ 0.095	1.196 $\pm$ 0.019	1.189 $\pm$ 0.016	1.206 $\pm$ 0.016
PD	1.128 $\pm$ 0.034	1.126 $\pm$ 0.037	1.121 $\pm$ 0.031	1.125 $\pm$ 0.045
PA	1.287 $\pm$ 0.095	1.257 $\pm$ 0.050	1.247 $\pm$ 0.039	1.285 $\pm$ 0.047

PBLUP= BLUP tradicional usando información de pedigrí; GBLUP: BLUP con información de marcadores; GBLUP+A: BLUP con información de marcadores y pedigrí; RKHS: *Reproducing Kernel Hilbert Space*.

Los coeficientes de regresión de los VGs predichos con un modelo sobre los VGs predichos con los demás modelos se muestran en el Cuadro 12. Estimadores de estos coeficientes igual a la unidad, indican que los VGs predichos con los diferentes modelos son similares (Mäntysaari, Liu & VanRaden, 2010). Estimadores de estos coeficientes más alejados de la unidad pueden interpretarse como sesgo de los VGs predichos con los modelos en comparación. Cuando se utilizó como variable dependiente los VGs predichos con PBLUP y como variable independiente los VGs predichos con cada uno de los modelos genómicos, los coeficientes de regresión estuvieron más alejados de la unidad para PN y PA, lo que quiere decir que los VGs predichos con los modelos genómicos tuvieron más sesgo con respecto a los predichos con PBLUP. Para los VGs correspondientes a PD, el sesgo fue menor, los estimadores variaron de 0.948 $\pm$ 0.013 a 0.968 $\pm$ 0.012. Las diferencias entre los coeficientes correspondientes a los modelos genómicos, considerando todas las variables analizadas, fueron pequeñas. Los resultados para este criterio de comparación de los modelos, son congruentes con los mencionados para habilidad de predicción y jerarquización de los VGs predichos, discutidos previamente.

Cuadro 12. Estimadores de los coeficientes de regresión ($\pm$ error estándar), utilizando como variable dependiente los valores genéticos predichos mediante un modelo y como variable independiente los valores predichos con los modelos restantes, entre paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, para pesos al nacimiento, al destete y al año en bovinos Suizo europeo.

Modelos	PBLUP	GBLUP	GBLUP+A	RKHS
PN (<0.0001)				
PBLUP		0.575 $\pm$ 0.028	0.569 $\pm$ 0.028	0.578 $\pm$ 0.028
GBLUP	1.356 $\pm$ 0.066		0.989 $\pm$ 0.004	1.004 $\pm$ 0.004
GBLUP+A	1.367 $\pm$ 0.067	1.008 $\pm$ 0.004		1.013 $\pm$ 0.003
RKHS	1.349 $\pm$ 0.066	0.994 $\pm$ 0.004	0.985 $\pm$ 0.003	
PD (<0.0001)				
PBLUP		0.968 $\pm$ 0.012	0.948 $\pm$ 0.013	0.963 $\pm$ 0.011
GBLUP	1.016 $\pm$ 0.013		0.976 $\pm$ 0.009	0.992 $\pm$ 0.005
GBLUP+A	1.034 $\pm$ 0.014	1.015 $\pm$ 0.009		1.011 $\pm$ 0.005
RKHS	1.023 $\pm$ 0.012	1.004 $\pm$ 0.005	0.985 $\pm$ 0.005	
PA (<0.0001)				
PBLUP		0.804 $\pm$ 0.031	0.773 $\pm$ 0.031	0.823 $\pm$ 0.028
GBLUP	1.089 $\pm$ 0.042		0.962 $\pm$ 0.010	1.007 $\pm$ 0.009
GBLUP+A	1.121 $\pm$ 0.045	1.028 $\pm$ 0.011		1.042 $\pm$ 0.009
RKHS	1.091 $\pm$ 0.038	0.985 $\pm$ 0.009	0.952 $\pm$ 0.008	

PBLUP= BLUP tradicional usando información de pedigrí; GBLUP: BLUP con información de marcadores; GBLUP+A: BLUP con información de marcadores y pedigrí; RKHS: *Reproducing Kernel Hilbert Space*; PN= Peso al nacimiento; PD = Peso al destete y PA= Peso al año.

Los resultados estimados en el presente estudio deberán tomarse con precaución dado que la población genotipada fue pequeña. Por lo que es importante considerar nuevamente la estimación de los efectos de los marcadores, e incrementa el número de animales genotipificados para obtener resultados más confiables.

3.6 Conclusiones

Las predicciones de valores genéticos para características de crecimiento en bovinos Suizo europeo pueden mejorarse con modelos de evaluación genética que utilizan información de marcadores SNP comparados con el modelo animal.

Los modelos de evaluación genética basados en marcadores SNP mejoran los estimadores de heredabilidad directa, la habilidad de predicción, las asociaciones de los fenotipos pre-ajustados con los valores genéticos predichos, y reducen el sesgo de los valores genéticos predichos para los pesos al nacimiento y al año en bovinos Suizo europeo comparadas con la metodológica cuantitativa tradicional.

La jerarquización de los valores genéticos predichos con los modelos basados en marcadores SNP fue similar que la correspondiente a los valores genéticos predichos con el modelo animal para peso al destete de bovinos Suizo europeo.

3.7 Literatura citada

- Bastiaansen, J. W. M., Coster, A., Calus, M. P. L., van Arendonk, J. A. M., & Bovenhuis, H. (2012). Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genetic Selection Evolution*, 44, 3.
- Buzanskas, M. E., Grossi, D.A., Ventura, R. V., Schenkel, F. V., Sargolzaei M., Meirelles S. L. C.,...Munari, D. P. (2004). Genome-wide association for growth traits in Canchim beef cattle. *Plos One*, 9(4), e94802.
- Calus, M. P., Huang, H., Vereijken, A., Visscher, J., Napel, J., & Windig, J. (2014). Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genetics Selection Evolution*, 1, 46-47.
- Christensen, O. F., Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42, 2-8.
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185, 1021-1031, doi: 10.1534/genetics.110.
- de los Campos, G., Naya H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K. & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182, 375-385.
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. (2013). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, 193, 327-345.
- Goddard, M. E., Hayes, B. J., & Meuwissen T. H. E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal Animal Breeding Genetics*, 128, 409-421. doi: 10.1111/j.1439-0388.2011.00964.
- Gonzalez-Recio, O., Gianola, D., Long, N., Weigel, K. A., & Rosa, G. J. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, 178, 2305-2313, doi:10.1534/genetics.107.084293.
- Habier, D., Fernando, R. L., & Dekkers, J. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177, 2389-2397. doi:10.1534/genetics.107.081190.
- Hayes, B., Bowman, P., Chamberlain, A., & Goddard, M. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, 92, 433-443.
- Kneeland, J., Li, C., Basarab, J., Snelling, W. M., Benkel, B., Murdoch, B., Hansen, C., & Moore, S. S. (2004). Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19,

- 21, and 23 within one commercial line of Bos B taurus. *Journal of Animal Science*, 82, 3405–3414.
- Mäntysaari, E., Zengting, L., & Van Raden, P. (2010). Interbull validation test for genomics evaluations. *Interbull Bolletín*, 41, 17-22.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157, 1819-1829.
- Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92, 4648-4655.
- Pérez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198, 483-495.
- Pszczola, M., Strabel, T., Mulder, H. A., & Calus, M. P. L. (2012). Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science*, 95(1), 389-400.
- Ríos-Utrera., A. (2008). Estimadores de parámetros genéticos para características de crecimiento predestete de bovinos. Revisión. *Técnica Pecuaria en México*, 46(1), 37-67.
- Roso, V. M., & Schenkel, F. S. (2006). AMC- A computer program to assess the degree of connectedness among contemporary groups. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, Brazil, August 13-18.
- Ryu, J. and Lee, C. (2014). Genomic Heritability of Bovine Growth Using a Mixed Model. *Journal Animal Science*, 27(11), 1521-1525.
- Saavedra-Jiménez, L. A., Rmírez-Valverde, R., Núñez-Domínguez, R., García-Muñiz, J. G., López-Villalobos, N., & Ruíz-Flores, A. (2013). Genotype by climate interaction in the gentic evaluation for growing traits of Braunvieh cattle in Mexico. *Tropical Anial Health and Production*, 45(7), 1489-1494. doi:10.1007/s11250-013-0387-2.
- SMN (Servicio Meteorológico Nacional). (2017). Información climatológica. Disponible en: <http://smn.cna.gob.mx/es/climatologia/informacion-climatologica>. Consultada el 30 de agosto de 2017.
- Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved Heritability Estimation from Genome-wide SNPs. *Animal Journal Human Genetics*, 91(6), 1011-1021.
- Snelling, W.M., Allan, M. F., Keele, J. W., Kuehn, L. A., McDanel, T., Sonstegard, T. P. L., Thallman, R. M., & Bennett, G. L. (2014). Genome-wide association study of growth in crossbred beef cattle. *Journal of Animal Science*, 88, 837-848.

- Tusell, L., Pérez-Rodríguez, P., Forni, S., Wu X. L., & Gianola, D. (2011) Genome-enabled methods for predicting litter size in pigs: a comparison. *Animal*, 7(11), 1739-1749. doi: 10.1017/S1751731113001389.
- VanRaden, P. M., Van Tasell, G. R., Wiggans, T. S., Sonstergard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24.
- VanRaden, P. M., & Sullivan, P. G. (2010). International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution*, 42, 7, doi:10.1186/1297-9686-42-7.
- Weigel, K. A., de los Campos, G., Gonzáles-Recio, O., Naya, H., Wu, X. L., Long, N., Rosa, G. J. M., & Gianola, D. (2009). Predictive ability of direct genomic values for lifetime net merit of Holstein sires using select subsets of single nucleotide polymorphism markers. *Journal of Dairy Science*, 23, 5248-5257. doi: 10.3168/jds.2009-2092.
- Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., . . . Dekkers, J. C. (2011). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution*, 43, 5.

3.8 Apéndices

Apéndice A. Rutina en R-3.4.3 para pre-ajuste de registros fenotípicos

```
#####  
#                               Ajuste de características                               #  
#####  
  
# Cargar las librerías y preparar los datos  
rm(list=ls())  
library(Matrix)  
library(lme4)  
library(XLConnectJars)  
library(XLConnect)  
setwd("C:/Users/Susana/Desktop/Archivos para tesis/Ajuste de  
características/Ajuste")  
SuizoE=read.csv(file="BW.csv", header=TRUE)  
  
# Estadísticos descriptivos para variable de interés  
length(SuizoE$BW); min(SuizoE$BW); max(SuizoE$BW);  
mean(SuizoE$BW); sd(SuizoE$BW)  
  
# Distribución de frecuencias para BW  
hist(SuizoE$BW,main="Distribución de frecuencias para BW",  
ylab="Frecuencia", xlab="BW(Birth Weigh (kg))", col="green")  
  
# Ajuste de peso al nacimiento  
AJUSTEBW1<-lmer(BW~(1|renCGBW), data=SuizoE)  
plot(SuizoE$BW,predict(AJUSTEBW1), main="Pre-ajuste en PN",  
ylab="Fenotipos predichos",xlab="Fenotipos observados",  
col="midnightblue")  
  
# Obtener valores predichos y residuales  
YHAT1<-predict(AJUSTEBW1); RESIDUALES<-SuizoE$BW-YHAT1  
  
# Exportar a excel valores predichos y residuales  
SuizoAjBW<-loadWorkbook("SuizoAjBW.xlsx", create=TRUE)  
createSheet(SuizoAjBW,name= "PREDICHOSBW")  
createSheet(SuizoAjBW,name= "RESIDUALESBW")  
createName(SuizoAjBW, name="PREDICHOSBW",  
formula="PREDICHOSBW!$A$1")  
createName(SuizoAjBW, name="RESIDUALESBW",  
formula="RESIDUALESBW!$A$1")  
writeNamedRegion(SuizoAjBW,YHAT1,name= "PREDICHOSBW")  
writeNamedRegion(SuizoAjBW,RESIDUALES,name= "RESIDUALESBW")  
saveWorkbook(SuizoAjBW)
```

```

# Ajuste de peso al destete
SuizoE=read.csv(file="WW.csv", header=TRUE)

# Estadísticos descriptivos para variables de interés
length(SuizoE$WW); min(SuizoE$WW); max(SuizoE$WW);
mean(SuizoE$WW); sd(SuizoE$WW)

# Obtener quantile para el 95% de información
Infquantile=read.csv(file="WW.csv", header=TRUE); AgeWW<-
Infquantile$AgeWW
quantile<-quantile(AgeWW,c(0.025,0.975))
#Base de datos con 95% información
SuizoQ=read.csv(file="WW-QUANTILE.csv", header=TRUE)

# Distribución de frecuencias para peso al destete
hist(SuizoQ$AgeWW,main="Distribución de frecuencias para WW",
ylab="Frecuencia", xlab="WW(Weaning Weight (kg))", col="green")

# Ajuste de peso al destete
AJUSTEWW1<-lmer(WW~ (1|renCGWW), data=SuizoQ)
plot(SuizoQ$WW,predict(AJUSTEWW1), main="Pre-ajuste en PD",
ylab="Fenotipos predichos",xlab="Fenotipos observados",
col="navy")

# Obtener valores predichos y residuales
YHAT1<-predict(AJUSTEWW1); RESIDUALES<-SuizoQ$WW-YHAT1

# Exportando a excel
SuizoAjWW<-loadWorkbook("SuizoAjWW.xlsx", create=TRUE)
createSheet(SuizoAjWW,name= "PREDICHOSWW1")
createSheet(SuizoAjWW,name= "RESIDUALESWW1")
createName(SuizoAjWW, name="PREDICHOSWW",
formula="PREDICHOSWW1!$A$1")
createName(SuizoAjWW, name="RESIDUALESWW1",
formula="RESIDUALESWW1!$A$1")
writeNamedRegion(SuizoAjWW,YHAT1,name= "PREDICHOSWW")
writeNamedRegion(SuizoAjWW,RESIDUALES,name= "RESIDUALESWW1")
saveWorkbook(SuizoAjWW)

# Ajuste de peso al año
SuizoE=read.csv(file="YW.csv", header=TRUE)

# Obtener un quantile para el 95% de información
AgeYW<-SuizoE$AgeYW; quantile<-quantile(AgeYW,c(0.025,0.975))
#Base de datos con 95% información
SuizoQ=read.csv(file="YW-QUANTILE.csv", header=TRUE)

```



```

# Distribución de frecuencias para peso al año
hist(SuizoQ$AgeYW,main="Distribución de frecuencias para WW",
ylab="Frecuencia", xlab="WW(Weaning Weigth (kg))", col="green")

# Ajuste peso al año
AJUSTEYW2<-lmer(YW~ (1|renCGYW), data=SuizoQ)
plot(SuizoQ$YW,predict(AJUSTEYW2), main="Pre-ajuste en PA",
ylab="Fenotipos predichos",xlab="Fenotipos observados",
col="royalblue4")
#obtener valores predicho y residuales
YHAT1<-predict(AJUSTEYW2)
RESIDUALES<-SuizoQ$YW-YHAT1

# Exportando a excel
SuizoAjYW<-loadWorkbook("SuizoAjYW.xlsx", create=TRUE)
createSheet(SuizoAjYW,name= "PREDICHOSYW1")
createSheet(SuizoAjYW,name= "RESIDUALESYW1")
createName(SuizoAjYW, name="PREDICHOSYW",
formula="PREDICHOSYW1!$A$1")
createName(SuizoAjYW, name="RESIDUALESYW1",
formula="RESIDUALESYW1!$A$1")
writeNamedRegion(SuizoAjYW,YHAT1,name="PREDICHOSYW")
writeNamedRegion(SuizoAjYW,RESIDUALES,name="RESIDUALESYW1")
saveWorkbook(SuizoAjYW)

```

Apéndice B. Rutinas en R-3.4.3 para construcción de matriz X, codificación, imputación y matriz G

```
#####  
#                               Construcción matriz X y G                               #  
#####  
  
# Cargar y preparar los datos  
rm(list=ls())  
setwd("C:/Users/Susana/Desktop/Archivos para tesis/Modelos  
genómicos/Generar matriz X")  
marcadores <- read.table("Marcadores .txt",skip=1)  
  
# Número de marcadores por animal  
size <- table(marcadores$V2); size <- unique(marcadores$V2); p  
<- size;  
n <- length(size)  
  
# Datos para construir la matriz de snp's  
dat <-  
data.frame(marcador=marcadores$V1,animal=marcadores$V2,alelo1=ma  
rcadores$V7,alelo2=marcadores$V8)  
rm(marcadores)  
  
# Función para recodificar  
snp.fun <- function(al1,al2){  
  if(al1=="-"&&al2=="-") snp <- NA  
  else if(al1!=al2) snp <- 1  
  else if(al1=="A") snp <- 0 else snp <- 2  
  return(snp)  
}  
snp.fun  
# snp's en un vector  
snps <-  
sapply(1:nrow(dat),function(i)snp.fun(dat$alelo1[i],dat$alelo2[i  
]))  
  
# snp's en matriz  
X1 <- matrix(snps,nrow=n,ncol=p,byrow=TRUE)  
dimnames(X1) <- list((size), unique(dat$marcador))  
  
# Información de marcadores y animales  
MARCADORES_INFO=X1  
MARCADORES_INFO
```

```

# Número de individuos
number_ind= rownames(MARCADORES_INFO); length(number_ind)

# Número de marcadores
number_marc=colnames(MARCADORES_INFO); length(number_marc)

# Marcadores con valores FALTANTES
MARCADORES_INFO[,28]
sum(is.na(MARCADORES_INFO[,28]))
tmp=is.na(MARCADORES_INFO)
sum(as.vector(colSums(tmp))==150)
FALTANTES<-which(as.vector(colSums(tmp))==150)
dim(MARCADORES_INFO)
X=MARCADORES_INFO[,-c(FALTANTES)]

# Recuperar matriz de marcadores
X=X

# Cargar librería para imputar marcadores
source("Impute.R")

# Imputación de marcadores
set.seed(123)
out=Impute(X)
str(out)

# Marcadores monomorphicos
monomorphic<-out$monomorphic
X[,114]
X[,29892]
X[,29201]

X=out$X[,-monomorphic]
dim(X)
is.na(X)
sum(is.na(X))

# Calcular P
phat=colMeans(X)/2
phat
MAF=ifelse(phat<0.5,phat,1-0.5)
MAF
phat=MAF
hist(MAF,main="MAF")
which(MAF<0.05)
MAF[which(MAF<0.05)]

```

```
# Eliminar marcadores con MAF menor que 0.05
X=X[,-which(MAF<0.05)]
X
fix(X)
dim(X)
rownames(X)
dim(X)

#Matriz de relaciones genómicas aditivas
X3=scale(X,center=TRUE,scale=TRUE)
G3=tcrossprod(X3)/ncol(X3)
G3
heatmap(G3)
```

Apéndice C. Rutina en R-3.4.3 para construcción de matriz A

```
#####  
# Matriz A de relaciones genéticas aditivas derivadas de un pedigrí #  
#####  
# Cargar y preparar los datos  
library(Matrix)  
library(lme4)  
library(pedigreeemm)  
setwd("C:/Users/Susana/Desktop/Archivos para tesis/Generar matriz  
A")  
dir()  
a=read.csv("Información pedigree.csv",header=TRUE)  
  
a=a[a[,2]!=0 & a[,3]!=0,]  
a=a[!duplicated(a),]  
  
pede=editPed(sire=a$Sire,dam=a$Dam,label=a$Animal,verbose=TRUE)  
ped=with(pede, pedigree(label=label, sire=sire, dam=dam))  
  
Xfull=relfactor(ped)  
dim(Xfull)  
str(Xfull)  
save(list=ls(all.names=TRUE),file="Xfull.RData",envir=  
.GlobalEnv)  
  
# Archivo con matriz Xfull  
load("C:/Users/Susana/Desktop/Archivos para tesis/Modelos  
genómicos/Generar matriz A/Xfull.RData")  
Xfull  
  
# Archivo con ids de animales con peso al nacimiento  
setwd("C:/Users/Susana/Desktop/Archivos para tesis/Generar  
matriz A")  
ids=read.csv("IDS_ANIMALSBW.csv",header=FALSE)  
str(ids)  
ids <- as.character(ids$V1)  
ids  
idsall<-rownames(Xfull)  
indice=idsall%in%ids  
X<-Xfull[,indice]  
dim(X)  
  
# Matriz A de relaciones genéticas aditivas  
A=as.matrix(t(X)%*%X)
```

Apéndice D. Rutina en R-3.4.3 para análisis de datos con modelo GBLUP o BLUP con marcadores

```
#####  
#      Modelo GBLUP o BLUP con marcadores para peso al nacimiento      #  
#####  
  
# Cargamos librerías  
library(Matrix)  
library(lme4)  
library(pedigreemm)  
library(BGLR)  
library(sommer)  
  
# Matriz X para individuos con PN  
idsall<-rownames(X)  
indice=idsall%in%ids  
X1<-X[indice,]  
dim(X1)  
  
# Matriz de relaciones genómicas para individuos con PN  
G<-G3  
ids=read.csv("IDS_BW.csv",header=FALSE)  
ids <- as.character(ids$V1)  
idsall<-rownames(G)  
indice=idsall%in%ids  
G1<-G[indice,indice]  
dim(G1)  
  
# Matriz A para individuos con PN  
A=as.matrix(t(X)%*%X)  
  
# Fenotipos para peso al nacimiento  
y=read.csv("FENOTIPOS BW.csv" ,header=TRUE)  
y  
if(any(rownames(G1)!=y$ID_INTERNO)) stop("Ordering problem\n")  
  
# Vector con fenotipos pre-ajustado  
y<-y[,8]  
  
# Matriz de incidencia para sexo  
y$SEX=as.factor(y$SEX)  
w1=model.matrix(~SEX-1,data=y)  
  
# MATriz de incidencia para edad de la madre lineal  
w2=model.matrix(~AM-1,data=y)
```

```

# Matriz de incidencia para edad de la madre cuadrático
w3=model.matrix(~AM2-1,data=y)

ETA=list(list(X=w1, model="FIXED"),
         list(X=w2, model="FIXED"),
         list(X=w3, model="FIXED"),
         list(K=G1, model="RKHS"))
fmGBW<-BGLR(y=y,ETA=ETA,nIter=10000,burnIn=5000,thin=10)
str(fmGBW)

# Predicciones
# Predicción total
yHat<-as.matrix(fmGBW$yHat)
SD.Hat<-fmGBW$SD.yHat

plot(fmGBW$yHat,y,xlab='Valores predichos', ylab='Fenotipos',
main='GBLUP',col= 2); cor(fmGBW$yHat,y)

# Varianza genética aditiva
varU<-fmGBW$ETA[[4]]$varU
varU

# Varianza residual
varE=scan("varE.dat")
varE

# Varian fenotípica
varP<- varU + varE

# Heredabilidad
h2<-varU/varP
h2

```

Apéndice E. Rutina en R-3.4.3 para análisis de datos con modelo GBLUP +A o BLUP con información de marcadores y pedigrí

```
#####
# Modelo GBLUP + A o BLUP con información de marcadores y pedigrí
para peso al nacimiento
#####

# Computing el eigenvalor de descomposición de g
EVD<-eigen(G1)

# Predictor lineal
ETA=list(list(X=w1, model="FIXED"),
         list(X=w2, model="FIXED"),
         list(X=w3, model="FIXED"),
         list(V=EVD$eigenvalues,
              d=EVD$values, model="RKHS"), list(K=A1, model="RKHS"))
fmGbp<-BGLR(y=y, ETA=ETA, nIter=10000, burnIn=5000, thin=10)
str(fmGbp)

# Predicciones
# Predicción total
yHat<-as.matrix(fmGbp$yHat)
SD.Hat<-as.matrix(fmGbp$SD.yHat)

plot(fmGbp$yHat, y[,8], xlab='Valores predichos',
     ylab='Fenotipos',
     main='GBLUP', col= 2)
cor(fmGbp$yHat, y)

# Varianza residual
varE=scan("varE.dat")

# Varianza genética aditiva de marcadores y pedigrí
varU1=fmGbp$ETA[[4]]$varU

varU2=fmGbp$ETA[[5]]$varU

# Heredabilidad
h2=(varU1+ varU2)/(varU1+ varU2+varE)
```


Apéndice F. Rutina en R-3.4.3 para análisis de datos con modelo RKHS

```
#####  
# Modelo RKHS con información de marcadores para peso al nacimiento #  
#####  
# Predictor lineal  
n<-nrow(X)  
p<-ncol(X)  
  
#Calculando D y K, establecer parametros h de 0 a 1  
X1<-scale(X, center=TRUE, scale=TRUE)  
D<-(as.matrix(dist(X1,method='euclidean'))^2)/p  
h<-0.5*c(1/5,1,5)  
  
ETA=list(list(X=w1, model="FIXED"),  
          list(X=w2, model="FIXED"),  
          list(X=w3, model="FIXED"),  
          list(K=exp(-h[1]*D),model='RKHS'),  
          list(K=exp(-h[2]*D),model='RKHS'),  
          list(K=exp(-h[3]*D),model='RKHS'))  
  
fmGr<-BGLR(y=y,ETA=ETA,nIter=10000,burnIn=5000,thin=10)  
str(fmGr)  
  
# Predicciones  
# Predicción total  
yHat<-as.matrix(fmGr$yHat)  
plot(fmGr$yHat,y[,6],xlab='Valores predichos',  
      ylab='Fenotipos',col= 2)  
cor(fmGr$yHat,y)  
  
# Varianza residual  
varE=scan("varE.dat")  
  
# Varianza genética aditiva  
varU1=fmGr$ETA[[4]]$varU  
  
varU2=fmGr$ETA[[5]]$varU  
  
varU3=fmGr$ETA[[6]]$varU  
  
varU<-mean(varU1+ varU2 + varU3)  
  
# Heredabilidad  
h2=(varU)/(varU + varE)
```

Apéndice G. Validación cruzada con 50 particiones aleatorias

```
# Validación cruzada con 50 particiones aleatorias 20% en prueba
y 80% en entrenamiento
correlations2=rep(NA,50)
mean2=rep(NA,5)
n=length(y)
testing=0.20
ntesting=round(testing*n)

for(fold in 1:50)
{
  yNA=y
  whichNa=sample(1:n,size=ntesting,replace=FALSE)
  yNA[whichNa]=NA
  fm=BGLR(y=yNA,ETA=ETA,nIter=10000,burnIn=5000,thin=10)
  unlink("*.dat")
  correlations2[fold]=cor(fm$yHat[whichNa],y[whichNa])
  mean2[fold]=mean((fm$yHat[whichNa]-y[whichNa])^2)
```