



UNIVERSIDAD AUTÓNOMA CHAPINGO
DEPARTAMENTO DE ENSEÑANZA, INVESTIGACIÓN
Y SERVICIO EN ZOOTECNIA
POSGRADO EN PRODUCCIÓN ANIMAL

APLICACIÓN DE ALGUNAS HERRAMIENTAS GENÓMICAS
EN BOVINOS SUIZO EUROPEO

TESIS

Que como requisito parcial
para obtener el grado de:

DOCTOR EN CIENCIAS EN INNOVACIÓN GANADERA

Presenta:

JONATHAN EMANUEL VALERIO HERNÁNDEZ

Bajo la supervisión de: **Agustín Ruíz Flores, Dr.**

Paulino Pérez Rodríguez, Dr.



Chapingo, Estado de México, mayo de 2022



APLICACIÓN DE ALGUNAS HERRAMIENTAS GENÓMICAS EN BOVINOS SUIZO EUROPEO

Tesis realizada por **Jonathan Emanuel Valerio Hernández**, bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

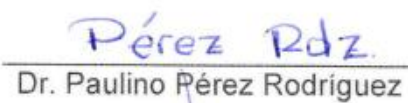
DOCTOR EN CIENCIAS EN INNOVACIÓN GANADERA

DIRECTOR:



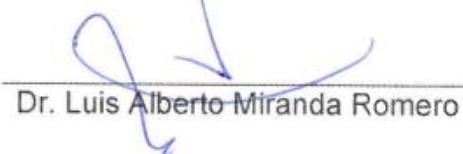
Dr. Agustín Ruíz Flores

CO-DIRECTOR:



Dr. Paulino Pérez Rodríguez

ASESOR:



Dr. Luis Alberto Miranda Romero

CONTENIDO

CONTENIDO	iii
LISTA DE CUADROS	v
LISTA DE FIGURAS	vii
LISTA DE APÉNDICES	viii
ABREVIATURAS USADAS	ix
DEDICATORIAS	x
AGRADECIMIENTOS	xi
DATOS BIOGRÁFICOS	xii
RESUMEN GENERAL	xiii
GENERAL ABSTRACT	xiv
1. INTRODUCCIÓN GENERAL	1
1.1. Literatura citada	3
2. REVISIÓN DE LITERATURA	5
2.1. Ciencias ómicas	5
2.1.1. Genómica	5
2.1.2. Transcriptómica	6
2.1.3. Proteómica	6
2.1.4. Metabolómica	7
2.1.5. Otras ómicas	9
2.2. Mejoramiento genético animal	10
2.2.1. Era de la selección genómica	12
2.2.2. Ventajas de la selección genómica	12

2.2.3.	Desventajas de la selección genómica.....	13
2.3.	Algunas metodologías estadísticas para la selección genómica...	14
2.3.1.	Regresión Ridge	16
2.3.2.	Regresión LASSO	17
2.3.3.	Regresión Elastic-Net	18
2.3.4.	Modelo de regresión GBLUP	19
2.3.5.	Modelo de regresión en un solo paso ssGBLUP	20
2.3.6.	Modelo de regresión cuantil QR	20
2.4.	Literatura citada.....	22
3.	SINGLE-STEP GENOMIC EVALUATION FOR GROWTH TRAITS IN A MEXICAN BRAUNVIEH CATTLE POPULATION	33
4.	REGRESIÓN CUANTIL PARA PREDICCIÓN DE CARACTERES COMPLEJOS EN BOVINOS SUIZO EUROPEO USANDO MARCADORES SNP Y PEDIGRÍ.....	54
5.	APÉNDICES	83

LISTA DE CUADROS

Cuadro	Capítulo	Título	Página
1	2	Listado de diferentes ómicas “establecidas” y “emergentes”	22
1	3	Estimates of additive genetic variance (σ_a^2), contemporary group variance (σ_{cg}^2), residual variance (σ_e^2) and heritability (h^2) for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a Mexican Braunvieh cattle population, using best linear unbiased predictor (BLUP), genomic BLUP (GBLUP) and single-step GBLUP (ssGBLUP) methods	49
2	3	Pearson correlation estimates between corrected phenotypes and estimated breeding values obtained from a validation with 100 random partitions when fitting (BLUP), genomic BLUP (GBLUP) and single-step GBLUP (ssGBLUP) models, for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a Mexican Braunvieh cattle population	51
1	4	Número de animales genotipados y fenotipados para el análisis de los pesos al nacer (PN), al destete (PD) y al año (PA) de una población de ganado Suizo Europeo.	74
2	4	Promedios de la correlación Pearson y la desviación estándar (entre paréntesis) entre los valores fenotípicos observados (y) y valores fenotípicos predichos ($\hat{y}$), cuadrado medio del error, componentes de varianza	75

asociados al error (σ_e^2 , σ_w^2) y criterio de información de desviación (DIC).

3	4	Promedios de la correlación Pearson y desviación estándar (entre paréntesis) entre efectos de marcador “verdaderos y estimados, señales “verdaderas” y estimadas, componentes de varianza asociados al error y valores de DIC para datos simulados con diferentes grados de asimetría y proporción de valores atípicos	78
---	---	--	----

LISTA DE FIGURAS

Figura	Capítulo	Título	Página
1	3	Distribution of the diagonal entries of a) Genomic relationship matrix derived from markers, G , b) the additive genomic relationship matrix derived from pedigree, A , and c) the combined matrix, H .	52
2	3	Pearson correlation coefficients between the observed and corrected phenotypes obtained with best linear unbiased predictor (BLUP), genomic BLUP (GBLUP), and single-step GBLUP (ssGBLUP) methods for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a Mexican Braunvieh cattle population.	53
1	4	Gráficas de violín de peso al nacimiento (PN), al destete (PD) y al año (PA) en una población de bovinos Suizo Europeo. La media está representada por el punto rojo y la mediana por la línea horizontal dentro del recuadro.	81

LISTA DE APÉNDICES

Apéndice 1. Código de programación en R para ajustar modelos de regresión cuantil.....	83
Apéndices 2. Código de muestreador de Gibbs en C, como complemento para ajustar modelos de regresión cuantil.....	90

ABREVIATURAS USADAS

ssGBLUP	Evaluación Genómica en un solo paso
BLUP	Mejor predictor lineal insesgado
GBLUP	Mejor predictor lineal insesgado genómico
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
QTL	Locus de rasgos cuantitativos
RR	Regresión Ridge
REN	Regresión Elastic-Net
LASSO	Regresión de operador de selección y contracción mínima absoluta
BLASSO	Regresión de operador de selección y contracción mínima absoluta bayesiana
SCE	Sumas de cuadrados del error
LARS	Algoritmo de regresión de ángulo mínimo
QR	Regresión cuantil
GEVB	Valores de cría genómicos
EVB	Valores de cría
BW, PN	Peso al nacimiento
WW, PD	Peso al destete
YW, PA	Peso al año
GWAS	Estudio de asociación del genoma completo
CG	Grupo contemporáneo
SNP	Polimorfismo de nucleótido simple
MCMC	Método de Cadenas de Markov Monte Carlo
MAF	Frecuencia del alelo menor
GC	Grupo contemporáneo
DIC	Criterio de información de desviación
MSE	Cuadrado medio del error

DEDICATORIAS

A nuestro Dios por permitirme entender cómo funciona un poco de su gran creación; como decía Louis Pasteur “La poca ciencia aleja de Dios, mientras que la mucha ciencia devuelve a Él”.

A mis profesores, por su paciencia, inspiración, conocimientos compartidos, ayuda y motivaciones para conmigo.

A mis padres por los consejos, motivaciones y ayuda durante mi estancia en Chapingo.

A la familia Tapia Barrera por abrirme la puerta de su casa, por los consejos, por el apoyo, atenciones para conmigo y considerarme parte de su familia.

A mis amigos y compañeros de clase, por los momentos de convivencia, por hacer amena, divertida e interesante mi vida.

A todas las personas que hicieron posible esto pero que no mencioné porque la lista sería muy grande.

¡GRACIAS A TODOS!

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología por el otorgamiento de la beca para realizar mis estudios de doctorado.

A la Universidad Autónoma Chapingo por recibirme y permitirme formar parte de su alma mater.

Al Dr. Agustín Ruíz Flores por su supervisión, consejos y motivaciones durante la maestría, doctorado y por su amistad.

Al Dr. Paulino Pérez Rodríguez por su supervisión, consejos y orientación en los análisis estadísticos del presente trabajo y por su amistad.

Al Dr. Luis Alberto Miranda Romero por las acertadas observaciones y recomendaciones.

A todos mis profesores por ayudar en mi formación profesional, ética y moral.

DATOS BIOGRÁFICOS



Datos personales

Nombre	Jonathan Emanuel Valerio Hernández
Fecha de nacimiento	24 de julio de 1993
Lugar de nacimiento	Sontecomapan, Catemaco, Veracruz
CURP	VAHJ930724HVZLRN07
Profesión	Ingeniero Agrónomo Zootecnista
Cédula Profesional	099880

Desarrollo académico

Doctorado en Ciencias (2018-2021)	DEIS en Zootecnia Posgrado en Producción Animal Universidad Autónoma Chapingo
Maestría en Ciencias (2016-2017)	DEIS en Zootecnia Posgrado en Producción Animal Universidad Autónoma Chapingo
Licenciatura (2008-2015)	DEIS en Zootecnia Universidad Autónoma Chapingo

RESUMEN GENERAL

APLICACIÓN DE ALGUNAS HERRAMIENTAS GENÓMICAS EN BOVINOS SUIZO EUROPEO¹

La raza bovina Suizo Europeo se utiliza en sistemas de producción de doble propósito del trópico mexicano por sus cualidades de adaptabilidad, fertilidad, producción de carne y leche. Si bien existen evaluaciones genéticas periódicas para esta raza, no se ha utilizado información genómica para complementar y mejorar la tasa de progreso genético anual. El objetivo de esta investigación fue utilizar información genómica para contribuir al mejoramiento genético de esta raza usando herramientas estadísticas y genómicas de vanguardia. En el primer estudio se utilizó la metodología estadística evaluación genómica en un solo paso (ssGBLUP), la que permite evaluar simultáneamente animales con y sin genotipo, se comparó con las metodologías mejor predictor lineal insesgado (BLUP) y genómico (GBLUP). La evaluación simultánea de animales con y sin genotipo mejoró la predicción de valores genéticos para las características de crecimiento, incluso con un número limitado de animales con genotipo. En un segundo estudio se utilizó la regresión cuantil (QR) como una alternativa para lidiar con distribuciones no normales y presencia de datos atípicos. QR se comparó bajo tres cuantiles (0.25, 0.50 y 0.75) contra GBLUP y ssGBLUP. Se encontró que el desempeño de QR es igual o superior a GBLUP y ssGBLUP, y que las ventajas de QR fueron más notorias con distribución asimétrica de los fenotipos y con mayor proporción de datos atípicos.

Palabras clave: Características de crecimiento, Suizo Europeo, Regresión cuantil, GBLUP, ssGBLUP.

¹Tesis de Doctorado en Ciencias en Innovación Ganadera, Posgrado en Producción Animal, Universidad Autónoma Chapingo

Autor: Jonathan Emanuel Valerio Hernández

Directores de Tesis: Agustín Ruíz Flores, Paulino Pérez Rodríguez

GENERAL ABSTRACT

APPLICATION OF SOME GENOMIC TOOLS IN BRAUNVIEH CATTLE²

The Braunvieh cattle breed is used in dual-purpose production systems in the Mexican tropics due to its qualities of adaptability, fertility, meat, and milk production. Although there are periodic genetic evaluations for this breed, genomic information has not been used to complement and improve the rate of annual genetic progress. The objective of this research was to use genomic information to contribute to the genetic improvement of this breed using some cutting-edge statistical and genomic tools. In the first study, the statistical methodology genomic evaluation in a single step (ssGBLUP) was used. It allows simultaneously evaluating animals with and without genotype, it was compared with the methodologies best linear unbiased (BLUP) and genomic (GBLUP) predictor. Simultaneous evaluation of genotyped and non-genotyped animals improved the prediction of growth traits, even with a limited number of genotyped animals. In the second study, quantile regression (QR) was used as an alternative to deal with non-normal distributions and the presence of outliers. QR under three quantiles (0.25, 0.50 and 0.75) was compared with GBLUP and ssGBLUP. It was found that the performance of QR is equal to or superior to GBLUP and ssGBLUP and that the advantages of QR were more noticeable with asymmetric distribution of phenotypes and with a higher proportion of outliers.

Key words: Growth traits, Braunvieh, Quantile Regression, GBLUP, ssGBLUP.

²Doctoral Thesis in Livestock innovation, Graduate Program in Animal Production, Universidad Autónoma Chapingo

Author: Jonathan Emanuel Valerio Hernández

Advisors: Agustín Ruíz Flores, Paulino Pérez Rodríguez

1. INTRODUCCIÓN GENERAL

La raza bovina Suizo Europeo se distribuye ampliamente en diferentes zonas de la República Mexicana. Estos animales se han introducido principalmente a las regiones tropicales y subtropicales de México; durante las últimas dos décadas ha estado reemplazando gradualmente al ganado *Bos indicus* debido a su adaptabilidad a ambientes hostiles, potencial lechero y características cárnicas que se adaptan muy bien a la demanda del mercado nacional (Arroyo-Rebollar, López-Villalobos, García-Martínez, Arriaga-Jordán, y Albarrán-Portillo, 2021). Las cualidades de esta raza la han convertido en una de las más importantes como fuente de material genético para hatos comerciales de doble propósito en el trópico mexicano, donde se utiliza como animales puros o cruza con ganado *Bos indicus* (Segura-Correa, Chin-Colli, Magaña-Monforte y Núñez-Domínguez, 2012).

Los primeros trabajos de mejoramiento genético del ganado Suizo Europeo en México se iniciaron en 2003 y desde entonces se ha hecho un seguimiento de la estabilidad de estas evaluaciones genéticas (Ramírez-Valverde, Núñez-Domínguez, Ruíz-Flores y García-Muñiz, 2008). Los esfuerzos del mejoramiento genético para esta raza han dado resultados. Recientemente, Larios-Sarabia et al. (2020) publicaron tendencias genéticas de algunas características de interés en bovinos Suizo Europeo, aunque de baja magnitud comparadas con las obtenidas antes de implementarse las evaluaciones genéticas. Dado el progreso genético limitado estos autores recomiendan replantear el programa nacional de mejoramiento genético para el bovino Suizo Europeo.

Recientemente, con el surgimiento de nuevas tecnologías genómicas, el acceso a los datos genómicos es cada día más común, haciendo factible el uso de esta información para mejorar la respuesta a la selección en animales (Hill, 2014; Meuwissen, Hayes y Goddard, 2016). Las posibilidades de uso de la información genómica son amplias y existen gran cantidad de trabajos que usan la información genómica con diferentes enfoques. Sin embargo, en el caso del ganado Suizo Europeo en México la información genómica no se ha usado ampliamente.

Recientemente, Valerio (2018) y Morales (2018) realizaron evaluaciones genómicas experimentales con información limitada de 300 animales genotipados.

Chin-Colli et al. (2015) calcularon parámetros genéticos de características de crecimiento de bovinos Suizo Europeo. Estos autores señalan que la selección para características de crecimiento puede dar lugar a respuestas favorables conjuntas, es decir, esta raza aún tiene potencial para ser mejorada y por ende provocar un impacto positivo en la producción en hatos comerciales; además se puede hacer uso de herramientas genómicas novedosas que combinadas con herramientas ya disponibles se espera den mejores resultados. Por lo anterior, el objetivo de la presente investigación es estudiar posibilidades para implementar nuevas metodologías que hagan uso de información genómica para la mejora genética del ganado Suizo Europeo.

En el Capítulo 2 de esta tesis, se revisa literatura acerca de avances en cuanto a ciencias ómicas, nuevas metodologías de evaluación genómicas y perspectivas de la selección genómica. En el Capítulo 3 se presenta un estudio donde se compara la evaluación genómica en un solo paso (ssGBLUP) contra la evaluación genómica (GBLUP) y las evaluaciones tradicionales (BLUP). En el Capítulo 4 se aborda el problema de evaluar datos con distribuciones sesgadas y presencia de valores atípicos; se implementa la regresión cuantil y se compara con análisis alternativos como GBLUP y ssGBLUP.

1.1. Literatura citada

- Arroyo-Rebollar, R., López-Villalobos, N., García-Martínez, A., Arriaga-Jordán, C. M., & Albarrán-Portillo, B. (2021). Relationship between calving interval and profitability of Brown Swiss cows in a subtropical region of Mexico. *Tropical Animal Health and Production*, 53(3), 373. <https://doi.org/10.1007/s11250-021-02813-0>
- Chin-Colli, R. del C., Estrada-León, R., Magaña-Monforte, J., Segura-Correa, J. C., & Núñez-Domínguez, R. (2015). Genetic parameters for growth and reproductive traits of brown swiss cattle from Mexico. *Ecosistemas y Recursos Agropecuarios*, 3(7), 11–20. Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-90282016000100002&nrm=iso
- Hill, W. G. (2014). Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. *Genetics*, 196(1), 1–16. <https://doi.org/10.1534/genetics.112.147850>
- Larios-Sarabia, N., Ramírez-Valverde, R., Núñez-Domínguez, R., García-Muñiz, J. G., & Ruiz-Flores, A. (2020). Impact of genetic evaluations on the genetic trends of Jersey and Brown Swiss cattle in Mexico. *Nova Scientia*, 12(24). <https://doi.org/10.21640/ns.v12i24.2305>
- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6(1), 6. <https://doi.org/10.2527/af.2016-0002>
- Morales, S. (2018). *Evaluación Genética de Bovinos Suizo Europeo Basada en Marcadores Moleculares*. Tesis de Maestría en Ciencias. Universidad Autónoma Chapingo.
- Ramírez-Valverde, R., Núñez, R., Ruíz-Flores, A., & García-Muñiz, J. G. (2008).

Stability of genetic evaluations of beef cattle Mexican populations. *Técnica Pecuaria En México*, 46(1), 13–24. Retrieved from https://www.academia.edu/5947823/Stability_of_genetic_evaluations_of_beef_cattle_Mexican_populations

Segura-Correa, J. C., Chin-Colli, R. C., Magaña-Monforte, J. G., & Núñez-Domínguez, R. (2012). Genetic parameters for birth weight, weaning weight and age at first calving in Brown Swiss cattle in Mexico. *Tropical Animal Health and Production*, 44(2), 337–341. <https://doi.org/10.1007/s11250-011-0026-8>

Valerio, J. E. (2018). *Comparación de Valores Genéticos Utilizando Marcadores SNP y Modelos Uni- y Bivariados*. Tesis de Maestría en Ciencias. Universidad Autónoma Chapingo.

2. REVISIÓN DE LITERATURA

2.1. Ciencias ómicas

Según Frigolet y Gutiérrez-Aguilar (2017) el término ómica fue acuñado por primera vez en los años 1980 y hace referencia a una amplia disciplina de la ciencia que estudia ciertas moléculas involucradas en el funcionamiento de un ser vivo. Dependiendo del conjunto de moléculas estudiadas estas ciencias reciben su prefijo, por ejemplo, si se estudian genes, genómica; transcriptos o ARNm, transcriptómica; proteínas, proteómica; o metabolitos, metabolómica. Existen otras que no han sido enlistadas ya que van surgiendo conforme aumenta la información disponible.

2.1.1. Genómica

La genómica es la ciencia que se encarga del estudio de los genomas, es decir el conjunto de información (ADN) de un organismo; tiene como objetivo predecir el origen de un fenotipo a partir de su genotipo (Fundación Instituto Roche, n.d.). La genómica surge con el proyecto del genoma humano (Venter et al., 2001) donde el objetivo principal era caracterizar todo el genoma; con el advenimiento de nuevas tecnologías, la secuenciación del genoma fue posible también en animales (Lippolis, Powell, Reinhardt, Thacker, y Casas, 2019). Los autores antes citados mencionan que en animales el principal interés es explicar el origen de fenotipos relacionados con salud y producción; para ello se han asociado ciertas regiones del ADN con rasgos de interés y reciben el nombre de QTL.

La genómica se divide en dos ramas principales: la genómica estructural orientada a la caracterización y localización de las secuencias que conforman el ADN, permitiendo de esta manera la obtención de mapas genéticos de los organismos; y la genómica funcional orientada a asignar las funciones correspondientes a los genes identificados, así como a las secuencias funcionales de ADN que no codifican para proteínas, con el fin de comprender el funcionamiento del genoma en su conjunto a través de los patrones de expresión (Fundación Instituto Roche, 2019).

2.1.2. Transcriptómica

La transcriptómica es la ciencia que se encarga del estudio de los transcriptos o ARNm presentes en una célula tejido u organismo bajo unas condiciones fisiológicas y ambientales concretas; es decir, estudia los perfiles de expresión de todos los genes en el genoma (Cartea, 2017; Kell y Oliver, 2016).

La tecnología de secuenciación ha permitido estudiar la expresión de todo el genoma, a esta expresión de genes se le conoce como transcriptómica, la cual tiene la capacidad no solo de cuantificar la expresión de la transcripción, sino también de estudiar las variaciones (Lippolis et al., 2019). En animales, la transcriptómica ha servido para establecer diferencias de expresión del ARN en características de interés como producción de leche (Dai, Zou, White, Liu, y Liu, 2018), características de crecimiento (Cánovas et al., 2010; Hao et al., 2020; Wiltafsky et al., 2010), reproductivas (Kropp et al., 2017; Wiltafsky et al., 2010), e inmunidad (Schroyen y Tuggle, 2015), entre otras.

2.1.3. Proteómica

La proteómica es la ciencia que estudia a escala masiva la distribución y número de proteínas codificadas por los genes, y sus interacciones expresadas en una célula tejido u órgano fijado en un momento dado, bajo determinadas condiciones y en una localización específica (Fundación Instituto Roche, 2019).

El estudio del transcriptoma se complementa con el análisis de las proteínas codificadas por los transcriptos; mientras que el genoma es invariable, el proteoma es dinámico; es decir, las proteínas cambian después de su síntesis dando como resultado que estas sean diferentes de una célula a otra, de un año a otro, e incluso de un momento a otro (Cartea, 2017). Este autor señala que el enfoque desde una perspectiva proteómica pretende estudiar cómo, dónde, cuándo y para qué están presentes las proteínas en un organismo y cómo interactúan entre ellas y con el ambiente.

Al igual que en la genómica hay subdivisiones, la proteómica, según la Fundación Instituto Roche (2019), se subdivide en:

La proteómica de expresión, que estudia de manera cualitativa y cuantitativa la expresión de proteínas, en términos de su abundancia relativa o de sus las proteínas con sus isoformas y modificaciones post-traduccionales, con el objetivo de detectar variaciones entre muestras que difieren en el tipo de proteínas expresadas o la cantidad en la que se expresa alguna de ellas; la proteómica estructural, dirigida a la caracterización de la estructura tridimensional de las proteínas; y la proteómica funcional, centrada en el estudio de la localización y distribución subcelular de proteínas y de las interacciones que existen entre ellas o con otras moléculas, con el fin de determinar su función.

La proteómica también se usa en animales con el enfoque principal de estudiar las proteínas relacionadas con rasgos de salud y producción como producción de leche (Le, Deeth, y Larsen, 2017; Wang et al., 2020), inmunidad (Euler, Hauck, Ueffing, y Deeg, 2013; Infantes-Lorenzo et al., 2017), crecimiento (Biancotto et al., 2019; Ma et al., 2020) y fertilidad (Q. Hao et al., 2019; Peddinti et al., 2008), entre otras.

2.1.4. Metabolómica

La metabolómica define como la ciencia enfocada al estudio de los metabolitos; su enfoque principal es medir y perfilar cambios en los niveles de metabolitos presentes dentro de una célula, tejido u organismo en respuesta a una variación genética o estímulos fisiopatológicos (Cartea, 2017; Park *et al.*, 2014).

Los metabolitos son pequeñas moléculas orgánicas presentes en los procesos celulares y muestran información del funcionamiento del metabolismo de una célula o ser vivo; el metabolismo se refiere al conjunto de reacciones químicas que ocurren en las células, tejidos u organismo (Cartea, 2017). La metabolómica permite analizar el perfil metabólico de una muestra tanto cuantitativa como cualitativa. La metodología es similar a las ómicas antes descritas y se busca encontrar

metabolitos específicos relacionados con alguna enfermedad o con la respuesta a un tratamiento nutricional o farmacológico (Frigolet y Gutiérrez-Aguilar, 2017). En el caso de animales se busca asociar metabolitos con características productivas de interés. La metabolómica tiene una gran ventaja comparada con las demás ciencias 'ómicas', esto porque el metaboloma es el producto final de la expresión génica, ofreciendo así una visión más realista de la función génica (Cartea, 2017). Sin embargo, el metaboloma contiene muchos tipos de moléculas biológicas con estructuras muy complejas, por lo que la metabolómica es mucho más compleja que las demás 'ciencias ómicas' (Kell y Oliver, 2016, citado por Cartea, 2017).

Dentro de la metabolómica no se existen divisiones como en proteómica o genómica, pero dado la gran cantidad de aportes en temas de obesidad y diabetes (Chevalier, Marliss, Morais, Lamarche, y Gougeon, 2005; Fiehn et al., 2010; Moore et al., 2014; Moran-Ramos et al., 2017), se considera dentro de la metabólica la lipidómica la cual Fundación Instituto Roche (2019) define de la manera siguiente:

La lipidómica es una rama de la metabolómica encargada de la caracterización molecular y funcional de los lípidos de un sistema biológico. La aplicación con más potencial de esta ciencia es la identificación de marcadores biológicos tanto de procesos patogénicos como de respuestas farmacológicas a distintas estrategias terapéuticas.

En animales los aportes de metabolómica están relacionados con características productivas de interés. Algunos ejemplos son: fertilidad (Kouassi et al., 2015; Uhde, Van Tol, Stout, y Roelen, 2018), producción de leche (Boudonck, Mitchell, Wulff, y Ryals, 2009; Huang et al., 2019), producción de carne (Stella et al., 2017; Yu, Tian, Shao, Li, y Dai, 2019), crecimiento (Kinkead et al., 2018), e inmunidad (Blakebrough-Hall, Dona, D'occhio, McMeniman, y González, 2020), entre otros.

2.1.5. Otras ciencias ómicas

Con el paso del tiempo nuevas tecnologías van surgiendo y más investigadores se suman aportando gran cantidad de información, lo que ocasiona que nuevas ciencias ómicas surjan para organizar mejor el conocimiento generado; la lista cada vez es más grande, algunas surgen de manera independiente y otras son productos de otras ómicas.

La Fundación Instituto Roche (2019) señala que existen tantas ciencias ómicas como elementos moleculares susceptibles a ser estudiados y en general se nombran añadiendo el sufijo “-ómica” al conjunto de moléculas estudiados, el Cuadro 1 adaptado de (Fundación Instituto Roche, 2019, p.12) muestran las ómicas ya consolidadas y algunas emergentes.

Cuadro 1. Listado de diferentes ómicas "establecidas" y "emergentes"

Ciencia ómica	Definición
Genómica	Estudio del conjunto del material genético presente en un organismo.
Farmacogenómica	Estudio de los genes que afectan a la respuesta de una persona a determinados fármacos.
Nutrigenómica	Estudio de la interacción entre genes y nutrientes.
Transcriptómica	Estudio de los perfiles de expresión de los ARN mensajeros, los microARNs y otros ARN no codificantes.
Epigenómica	Estudio de los elementos que controlan la expresión génica, sin modificar la secuencia de nucleótidos del ADN.
Proteómica	Estudio del conjunto completo de proteínas expresadas en un organismo en un tiempo determinado y particular de cada tipo celular o tisular.

Metabolómica	Identificación y cuantificación de productos metabólicos de pequeño tamaño (metabolitos) de un sistema biológico (célula, tejido, fluido biológico u órgano).
Lipidómica	Estudio de la estructura y función del conjunto de moléculas lipídicas en una célula u organismo, así como de sus interacciones con otros lípidos, proteínas o metabolitos.
Secretómica	Estudio de las moléculas orgánicas y compuestos inorgánicos secretados por las células, tejidos u órganos.
Interactómica	Estudio de las interacciones moleculares en un sistema biológico.
Citómica	Estudio de la colección de procesos celulares complejos y dinámicos subyacentes a los procesos fisiológicos, así como de la diversidad funcional y estructural de conjunto de células de un organismo.
Fenómica	Estudio del conjunto de los fenotipos (descripciones físicas) expresados por una célula, resultado de la interacción del genotipo y el ambiente.
Exposómica	Estudio del conjunto de factores ambientales a los que se expone una persona a lo largo de su vida.
Metagenómica	Estudio del conjunto de microorganismo de una muestra ambiental para proporcionar información de la diversidad ecológica de un ambiente determinado.
Microbiómica	Estudio del material genético de los microorganismos en un nicho específico, como el tracto gastrointestinal, la cavidad bucal o la piel.

Nota. Fuente: Fundación Instituto Roche, (2019).

2.2. Mejoramiento genético animal

La crianza de animales ha cambiado mucho a lo largo del tiempo. No se sabe el momento exacto en el que el humano comenzó a seleccionar animales para su

propósito y beneficio, pero probablemente esto ocurrió desde la domesticación de las especies. En sus inicios solo se seleccionaban aquellos que a la vista del criador mejor le pareciera. Picardi (2017) cita el libro de Bianchi (2014), el cual analiza desde el punto de vista científico una de las narrativas más antiguas de la humanidad, hace referencia al antiguo testamento y la historia de Jacob. Se puede observar que en esta narrativa están intrínsecos algunos conceptos básicos de herencia mendeliana, es decir, una de las evidencias más antiguas de selección en animales data del siglo VI A. C. la cual es la fecha en la que los eruditos en teología creen que fueron escritos los libros del pentateuco.

“Para poder independizarse de su suegro Labán y cumplir con el trato que éste le fijara, Jacob hace uso de los conocimientos que tenía sobre el pelaje con caracteres dominantes y recesivos en ovejas y cabras. Este trato, para que Jacob pudiera llevarse a su familia, consistía en devolver en siete años una cierta cantidad de ovejas negras y cabras moteadas. Labán, astutamente, había eliminado todos estos animales al hacer el trato. Pero Jacob ya había usado la metodología de la observación y sabía que las frecuencias de estos pelajes iban igualmente a presentarse. Con observación y constancia Jacob modificó estas poblaciones en su frecuencia de animales para su beneficio y así desligarse de Labán” (Picardi, 2017, p.43-44).

Otro criador que se basó en la observación para seleccionar animales fue Robert Bakewell (1725-1795), un inglés que fue reconocido como uno de los criadores más exitosos al definir razas de ovinos, bovinos y equinos (Iowa State University, n.d.; Picardi, 2017). Más tarde Galton (1822-1911) y Pearson (1857-1936) trabajaron con registros fenotípicos y encontraron relaciones entre padres e hijos, aún y cuando las leyes de Mendel no se habían descubierto (Gianola y Rosa, 2015).

Una aportación importante al mejoramiento genético animal fue realizado por Hazel y Lush (1942) ellos propusieron usar distintas fuentes de información (registros fenotípicos) de parientes para estimar el mérito genético del animal, de esta forma se establece por primera vez una metodología científica para la selección genética. Más tarde uno de los discípulos de Lush, Henderson (1953, 1963, 1975, 1984) hizo

gran variedad de aportes siendo la aplicación de los modelos lineales mixtos uno de los más importantes y vigente a la fecha (Tirados, 2001).

2.2.1. Era de la selección genómica

Los adelantos tecnológicos en biología molecular permitió acceder al genotipo de los animales, esta información comenzó a utilizarse en los procedimientos de evaluación genética dando paso a la selección asistida por marcadores moleculares (Fernando y Grossman, 1989). En un principio la genética molecular se enfocaba a entender la naturaleza y funcionamiento del gen, fue así como se identificaron algunos genes con efectos importantes en características estudiadas QTL (Quantitative Trait Loci) (Martínez, Manrique, y Elzo, 2012). La selección genómica se considera un tipo de selección asistida por marcadores donde se distribuyen gran cantidad de marcadores (por ejemplo SNPs) a lo largo del todo el genoma de tal manera que se espera que todos los loci de características cuantitativas queden en desequilibrio de ligamiento con al menos un marcador (Hayes et al., 2007). Los primeros en sugerir esta metodología fueron Meuwissen, Hayes y Goddard (2001), la llamaron selección genómica y la idea principal fue usar altas densidades de marcadores para después estimar el efecto de cada marcador, explicar la varianza genética y predecir los valores genéticos de los animales.

2.2.2. Ventajas de la selección genómica

Las posibilidades y ventajas de la selección genómica son amplias. Crossa et al. (2010) realizaron validaciones cruzadas en trigo con información de marcadores moleculares y pedigrí. Estos autores observaron que los modelos que incluyen información genómica son más precisos que los modelos basados únicamente en pedigrí, entre sus hallazgos reportaron ganancias de 7.7 a 35.7% en capacidad predictiva.

Las evaluaciones genéticas tradicionales necesitan que los animales tengan registros fenotípicos o en su defecto cuenten con familiares con registros, esto solo

se logra cuando el animal crece y expresa su potencial genético, lo cual puede tardar incluso años dependiendo de la especie. La disponibilidad de tecnologías moleculares actuales posibilita acceder al genotipo del animal a edades tempranas por lo que el intervalo generacional se acorta permitiendo así un incremento en la tasa de mejoramiento anual (Yudin, Lukyanov, Voevoda, y Kolchanov, 2016). La disminución del intervalo generacional puede representar ahorros importantes no solo de tiempo sino de dinero, se evitan costos de alimentación, mano de obra, instalaciones etc. Schaeffer (2006) señala que en bovinos lecheros se pueden seleccionar animales jóvenes mediante marcadores y evitar pruebas de progenie, lo que podría reducir los costos de operación hasta 92%.

Una forma en la que se realiza las evaluaciones genómicas es utilizar cierta población para entrenar modelos y obtener ecuaciones de predicción, esta cualidad puede aprovecharse para evaluar características difíciles de medir por ejemplo calidad de la canal, resistencia a enfermedades y vida productiva, entre otras. La cualidad de estas características hace que sea difícil la selección ya que por lo regular se necesita que el animal esté muerto o muy viejo y dificultándose así pasar el conjunto de genes deseado a la siguiente generación. La selección genómica puede resolver esta problemática ya que se puede usar ganado comercial para obtener ecuaciones de predicción y se pueden seleccionar animales jóvenes solo con su información genómica, es decir se puede obtener una buena predicción del valor genético del animal sin la necesidad de registros fenotípicos (Goddard y Hayes, 2007).

2.2.3. Desventajas de la selección genómica

Boichard, Ducrocq, Croiseau y Fritz (2016) señalan que la eficiencia de la selección genómica depende de muchos factores; por ejemplo, se requieren poblaciones que cuenten con registros fenotípicos y que estén conectados genéticamente entre ellos. A partir de estas poblaciones se puede conocer el tamaño y tipo de chip adecuado para la característica de interés, la correlación con otras características de interés,

el método estadístico apropiado y la habilidad predictiva de estos, entre otros. Sin embargo, muchas veces estas poblaciones de referencia no están disponibles, son pequeñas e insuficientes por lo costoso y el trabajo de años que implica para lograr consolidar una buena población de referencia.

La selección genómica usa gran cantidad de información de marcadores moleculares; con las tecnologías de genotipado modernas los paneles de marcadores disponibles son densos, esto implica que el número de efectos de marcador y parámetros a estimar puede superar considerablemente el número de registros disponibles. Para enfrentar estos inconvenientes de estimación de p grandes (número de marcadores) y n pequeños (número de registros fenotípicos) es necesario utilizar metodologías estadísticas más sofisticadas (de los Campos, Hickey, Pong-Wong, Daetwyler, y Calus, 2013; Jannink, Lorenz, e Iwata, 2010).

Goddard y Hayes (2007) mencionan que una de las principales limitantes para la implementación de la selección genómica es el costo del genotipado, actualmente el costo del genotipado ha disminuido y la tendencia es que se reduzca más (De Donato, Peters, Mitchell, Hussain, y Imumorin, 2013). Sin embargo, aún y con el costo actual del genotipado, sigue siendo una limitante para implementar la selección genómica (Rajsic, Weersink, Navabi, y Peter Pauls, 2016). Además, esto puede originar dependencia tecnológica, ya que la genotipificación se realiza con equipo especializado, generalmente costoso, técnicas especializadas y se tiene que realizar periódicamente (Goddard y Hayes, 2007). En algunos casos implementar la selección genómica no resulta tan conveniente, por ejemplo, animales adultos probados a través de su progenie suelen tener exactitudes de predicción altas por lo que la ganancia adicional en exactitud por incluir información genómica puede ser de baja magnitud (Goddard y Hayes, 2007; Rajsic et al., 2016).

2.3. Algunas metodologías estadísticas para la selección genómica

La idea principal de la selección genómica es distribuir gran cantidad de marcadores por todo el genoma de tal forma que se espera que queden en desequilibrio de

ligamientos con genes que afectan la característica de interés (Meuwissen et al., 2001). El problema de trabajar con gran cantidad de marcadores se le conoce en la literatura como “la maldición de la dimensionalidad”, que surge cuando el número de variables p (marcadores) es mayor que el número de observaciones n (individuos). Según Crossa (2010) existen otros desafíos a superar como la colinealidad entre marcadores o la complejidad de las características cuantitativas. El resultado es un sobreajuste que disminuye la capacidad predictiva del modelo para individuos no incluidos en la población de entrenamiento.

En los modelos genéticos estándar, los resultados fenotípicos, y_i ($i = 1, \dots, n$), se consideran como la suma de un valor genético, g_i , y un residual, ε_i ; es decir $y_i = g_i + \varepsilon_i$, (Crossa et al., 2010). En los modelos paramétricos para selección genómica, g_i se describe como una regresión en covariables de marcadores x_{ij} ($j = 1, \dots, p$, marcadores moleculares), de forma tal que $g_i = \sum_{j=1}^p x_{ij}\beta_j$, tal que $y_i = \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i$ (o en notación matricial, $y = X\beta + \varepsilon$), donde β_j es la regresión de y_i en la covariable correspondiente al j -ésimo marcador x_{ij} (Crossa et al., 2010; Meuwissen et al., 2001).

La estimación de los coeficientes de regresión asociados con los marcadores, β , mediante regresión múltiple por métodos convencionales, como máxima verosimilitud y mínimos cuadrados no son viables. Esto es porque el número de observaciones fenotípicas, n , es mucho menor que el número de marcadores, p , ($n \ll p$) y la inversa de X^tX puede no existir por la multicolinealidad entre marcadores (Crossa et al., 2010; Nicolás, 2013).

Algunas metodologías estadísticas se desarrollaron para calcular los efectos de los marcadores conocidos en la literatura como métodos penalizados tales como regresión Ridge (RR), LASSO, mejor predictor lineal insesgado genómico (GBLUP), la regresión elastic-net (REN), entre otros (Schaefer, Roi y Wolfe, 1984; López, 2012; de los Campos et al., 2013; Jacquin, Cao y Ahmadi, 2016). Estos métodos penalizados se basan en la minimización de la suma de cuadrados del error (SCE),

a través de restricciones de los posibles valores de los estimadores para reducir la varianza, obteniendo así estimaciones más precisas (Kyung, Gilly, Ghoshz y Casellax, 2010; Nicolás, 2013).

2.3.1. Regresión Ridge

La regresión Ridge (RR) es uno de los métodos de estimación penalizada más antiguos (Hoerl y Kennard, 1970). La RR estima simultáneamente el efecto de cada marcador, el modelo es el siguiente:

$$y_i = \sum_{j=1}^p x_{ij} \beta_j + \varepsilon_i$$

donde y_i es el vector de fenotipos, x_{ij} es un elemento de la matriz de incidencia que relaciona al animal con el efecto del marcador y ε_i es el vector de residuales.

En RR se considera el efecto de los marcadores como fijos y los residuales con distribución normal con media 0 y varianza constante σ_ε^2 . Los estimadores Ridge de los coeficientes de regresión se obtienen resolviendo la siguiente ecuación:

$$\hat{\beta}_{RR} = (\mathbf{X}^t \mathbf{X} + \lambda_{RR} \mathbf{I})^{-1} \mathbf{X}^t \mathbf{y}$$

donde $\lambda_{RR} \geq 0$ es una constante no negativa que controla la bondad de ajuste y la complejidad del modelo, medida por la suma de cuadrados de los coeficientes de regresión o la norma L_2 de β (López, 2012). La norma L_2 es equivalente a la distancia Euclídea del vector al centro de coordenadas, por lo que la penalización es proporcional a los cuadrados de los coeficientes. El parámetro de Ridge λ_{RR} , se debe estimar usando los datos (Le Cessie y Van Houwelingen, 1992; Lee y Silvapulle, 1988; Schaefer et al., 1984):

1. λ_0 = el valor de β_R donde se estabiliza el gráfico, conocido como la traza de Ridge.
2. $\lambda_1 = \frac{1}{\beta^t \beta}$

$$3. \lambda_2 = \frac{p+1}{\boldsymbol{\beta}^t \boldsymbol{\beta}}$$

$$4. \lambda_3 = \frac{(\lambda_1 - 100\lambda_p)}{99}.$$

donde λ_1 y λ_p son el menor y mayor valor propio de $\mathbf{X}^t \mathbf{X}$.

La RR presenta algunos inconvenientes, por ejemplo, no proporciona un modelo parsimonioso, ya que mantiene todas las variables en el modelo, es decir, el grado de contracción es homogéneo entre los marcadores (Zou y Hastie, 2005). La contracción homogénea limita una correcta estimación del efecto de cada marcador ya que algunos marcadores pueden estar localizados en regiones no asociados con la varianza genética y otros si (Goddard y Hayes, 2007). Otra desventaja de este método es que λ_{RR} depende de $\boldsymbol{\beta}$, que es desconocido y que penaliza a todos los β_j , estén o no en la relación de colinealidad (Nicolás, 2013).

2.3.2. Regresión LASSO

Tibshirani (1996) fue el primero en proponer la regresión LASSO (Least Absolute Shrinkage and Selection Operator) operador de selección y contracción mínima absoluta. Este tipo de regresión al igual que la regresión Ridge, es un caso particular de la regresión Bridge. La diferencia entre LASSO y RR es que RR hace las estimaciones a través de mínimos cuadrados con la restricción L_1 (Park y Casella, 2008), es decir los coeficientes de regresión lineal se obtienen al resolver la siguiente ecuación:

$$\hat{\beta}_L = \min_{\boldsymbol{\beta}} \left\{ (y - \mathbf{X}\boldsymbol{\beta})^t (y - \mathbf{X}\boldsymbol{\beta}) + \lambda_L \sum_{j=1}^p |\beta_j| \right\}$$

Para resolver esta ecuación se han propuesto varias soluciones una de las más usadas es el algoritmo LARS (Least Angle Regression) desarrollado por Efron, Hastie, Johnstone, y Tibshirani (2004).

La regresión LASSO selecciona variables y las penaliza simultáneamente haciendo algunas igual a cero por medio del parámetro regulador λ_L que se obtiene por validación cruzada (Bühlmann y van de Geer, 2011). La regresión LASSO a diferencia de RR tiene un modelo más parsimonioso ya que algunos coeficientes los hace cero, es decir realiza selección de variables y contracción simultáneamente (López, 2012; Nicolás, 2013).

La regresión LASSO presenta algunos inconvenientes por ejemplo siendo $p > n$, la regresión LASSO selecciona como máximo n variables; si existen altas correlaciones por pares entre marcadores, se elige uno de cada subconjunto sin importar cuál de estos es seleccionado, entonces si todos los predictores están correlacionados, la regresión LASSO se desempeña insatisfactoriamente y es superada por la RR (López Cruz, 2012; Nicolás, 2013; Tibshirani, 1996; Zou y Hastie, 2005).

2.3.3. Regresión Elastic-Net

La regresión Elastic-Net (REN) fue propuesta por Zou y Hastie (2005) con la intención de superar las desventajas de LASSO y RR, entonces REN corresponde a una generalización de LASSO y RR. La regresión REN ocupa las penalizaciones L_1 tipo LASSO y L_2 tipo RR (Bühlmann y van de Geer, 2011). Entonces el estimador de REN se obtiene minimizando la suma de cuadrados ordinarios con ambas restricciones la ecuación se muestra a continuación:

$$\hat{\beta}_{EN} = \min_{\beta} \left\{ (y - X\beta)^t (y - X\beta) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

donde λ_1 y λ_2 son constantes no negativas, las cuales controlan el peso asignado a las penalidades L_1 y L_2 , respectivamente. Este problema de optimización es equivalente a:

$$\min_{\beta} \{ (y - X\beta)^t (y - X\beta) \}, \text{ sujeto a las restricciones: } (1 - \alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \leq t$$

donde $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$ y t es una constante arbitraria positiva (López, 2012; Nicolás, 2013).

2.3.4. Modelo de regresión GBLUP

El modelo GBLUP (Genomic Best Linear Unbiased Prediction) es una modificación del método BLUP (Best Linear Unbiased Prediction) desarrollado por Henderson (1984). La única diferencia es que en GBLUP se reemplaza la matriz de relaciones genéticas **A** por la matriz de relaciones genómicas **G** derivada de la información de los marcadores moleculares (Clark y van der Werf, 2013).

Las predicciones de GBLUP suelen ser más precisas que las de BLUP ya que las relaciones genómicas suelen ser más exactas por ejemplo en una relación genética de medios hermanos pueden compartir 60% o 40% de sus alelos; con información derivada del pedigrí se asume que comparten el 50% de sus alelos, mientras que con información genómica se puede conocer la relación exacta entre los medios hermanos, he aquí la mayor precisión de estimadores GBLUP frente a aquellos obtenidos con BLUP (de los Campos et al., 2009).

El modelo estadístico para GBLUP es:

$$y = X\beta + Zu + \varepsilon$$

donde y es el vector de fenotipos, X es una matriz que relaciona los fenotipos con los efectos fijos, Z es una matriz de incidencia que relaciona el efecto aleatorio del animal con los fenotipos, β es el vector de efectos fijos y u es el vector de efectos aleatorios del animal. Las suposiciones para GBLUP son que $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$ y que $u \sim N(0, \sigma_u^2 G)$, G es la matriz de relaciones genómicas derivada de marcadores moleculares.

2.3.5. Modelo de regresión en un solo paso ssGBLUP

El modelo ssGBLUP es equivalente al modelo GBLUP descrito anteriormente con la diferencia de que se reemplaza la matriz de relaciones genómicas $\mathbf{G}$ por la matriz de relaciones genéticas extendida $\mathbf{H}$, se asume que $Var(\mathbf{u}) = \sigma_H^2 \mathbf{H}$.

La matriz $\mathbf{H}$ se construye con matrices de relaciones para animales genotipados y no genotipados y de los cuales se dispone un pedigrí. La matriz que resulta de acuerdo con Christensen y Lund (2010) y Aguilar et al. (2010), es:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_a^{-1} - \mathbf{A}_{gg}^{-1} \end{bmatrix},$$

donde, $\mathbf{A}_{gg}$ es una submatriz de $\mathbf{A}$ para animales genotipados, $\mathbf{G}_a = \beta \mathbf{G} + \alpha$; β y α se obtienen al resolver el sistema de ecuaciones:

$$\begin{cases} Avg(diag(\mathbf{G}))\beta + \alpha = Avg(diag(\mathbf{A}_{gg})) \\ Avg(\mathbf{G})\beta + \alpha = Avg(\mathbf{A}_{gg}) \end{cases}.$$

2.3.6. Modelo de regresión cuantil QR

Koenker y Bassett (1978) fueron los primeros en proponer la regresión cuantil (QR). Uno de los primeros trabajos donde se utiliza QR para predecir el mérito genético individual fue presentado por Nascimento y colaboradores (2017) quienes trabajaron con datos simulados y encontraron ventajas al usar QR frente a metodologías convencionales.

Unas de las características que hace que QR sea muy atractivo para los genetistas es la solidez de los supuestos de distribución y las observaciones atípicas; QR puede dar una visión amplia de los datos al ir probando la relación de diferentes cuantiles de la variable de respuesta y covariables, de tal forma que se pueden estudiar las colas de distribuciones; el uso de un enfoque sólido como QR hace que

la inferencia sea menos sesgada y esté menos sujeta a falsos positivos (Briollais y Durrieu, 2014).

El modelo para regresión cuantil está dado por:

$$y_i = \mu + \mathbf{x}_i^t \boldsymbol{\beta} + w_i,$$

donde y_i es el valor del fenotipo del i -ésimo animal; μ es un intercepto; $\mathbf{x}_i^t = (x_{i1}, \dots, x_{ip})$ representa la fila i -ésima de la matriz de marcadores, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ es el vector de coeficientes de regresión asociados a marcadores y w_i son variables aleatorias independientes tales que su cuantil $\theta \in (0,1)$ es cero. La estimación de los coeficientes de regresión para un cuantil de interés θ fijo se obtiene al resolver el problema de minimización siguiente:

$$\min\{\sum_{i=1}^n \rho_{\theta}(y_i - \mu - \mathbf{x}_i^t \boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j|\},$$

donde $\sum_{j=1}^p |\beta_j|$ es la suma de los valores absolutos de los coeficientes de regresión; λ es el parámetro de penalización que controla la intensidad de la regularización; y $\rho_{\theta}(\cdot)$ es la función definida como (Koenker y Bassett, 1978):

$$\rho_{\theta}(t_i) = \begin{cases} \tau \times t_i & \text{Si } t_i \geq 0 \\ -(1 - \tau) \times t_i & \text{Si } t_i < 0, \end{cases}$$

donde $t_i = y_i - \mu - \mathbf{x}_i^t \boldsymbol{\beta}$. Después de estimar los parámetros del modelo, los valores de cría predichos mediante marcadores (GEVB, Genomic Estimated Breeding Values) se obtienen mediante la siguiente expresión $GEBV(\tau) = \hat{y}_i(\tau) = \sum_{j=1}^p x_{ij} \hat{\beta}_j(\tau)$,

donde $\hat{\beta}_j(\tau)$ es el efecto del j -ésimo marcador, definido por la relación funcional obtenida para el cuantil de interés (Pérez-Rodríguez, Montesinos-López, Montesinos-López, y Crossa, 2020).

2.4. Literatura citada

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, 93(2), 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Bianchi, N. O. (2014). *Genética y arte*. 1a ed. (P. Blesa, Ed.). La plata, Argentina.
- Biancotto, G., Bovo, D., Mastroilli, E., Manuali, E., Angeletti, R., & Stella, R. (2019). TMT-based proteomics profiling of bovine liver underscores protein markers of anabolic treatments. *Proteomics*, 19(9). <https://doi.org/10.1002/pmic.201800422>
- Blakebrough-Hall, C., Dona, A., D'occhio, M. J., McMeniman, J., & González, L. A. (2020). Diagnosis of Bovine Respiratory Disease in feedlot cattle using blood ¹H NMR metabolomics. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-019-56809-w>
- Boichard, D., Ducrocq, V., Croiseau, P., & Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies*, 339(7-8), 274-277. <https://doi.org/10.1016/j.crvi.2016.04.007>
- Boudonck, K. J., Mitchell, M. W., Wulff, J., & Ryals, J. A. (2009). Characterization of the biochemical variability of bovine milk using metabolomics. *Metabolomics*, 5(4), 375–386. <https://doi.org/10.1007/s11306-009-0160-8>
- Briollais, L., & Durrieu, G. (2014). Application of quantile regression to recent genetic and -omic studies. *Human Genetics*, 133(8), 951–966. <https://doi.org/10.1007/s00439-014-1440-6>
- Bühlmann, P., & van de Geer, S. (2011). *Statistics for High-Dimensional Data*. 1st

ed. Heidelberg: Springer, Berlin, Heidelberg.
<https://doi.org/https://doi.org/10.1007/978-3-642-20192-9>

- Cánovas, A., Quintanilla, R., Amills, M., & Pena, R. N. (2010). Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics*, 11(1). <https://doi.org/10.1186/1471-2164-11-372>
- Cartea, Ma. E. (2017). Sociedad de Ciencias de Galicia. *Mol*, 17, 66–73. Retrieved from <http://mol-en.scg.org.es/wp-content/uploads/2018/01/MOL-17.pdf#page=66>
- Chevalier, S., Marliss, E. B., Morais, J. A., Lamarche, M., & Gougeon, R. (2005). Whole-body protein anabolic response is resistant to the action of insulin in obese women. *American Journal of Clinical Nutrition*, 82(2), 355–365. <https://doi.org/10.1093/ajcn.82.2.355>
- Christensen, O., & Lund, M. (2010). Genomic relationship matrix when some animals are not genotyped. *Genetics Selection Evolution*, 42(2), 1–8. Retrieved from <http://www.gsejournal.org/content/42/1/2>
- Crossa, J., L, De Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., ... Braun, H. J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2), 713–724. <https://doi.org/10.1534/genetics.110.118521>
- Dai, W. ting, Zou, Y. xuan, White, R. R., Liu, J. xin, & Liu, H. yun. (2018). Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Functional and Integrative Genomics*, 18(2), 125-140. <https://doi.org/10.1007/s10142-017-0580-x>
- De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. *PLoS ONE*,

8(5). <https://doi.org/10.1371/journal.pone.0062137>

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2), 327–345. <https://doi.org/10.1534/genetics.112.143313>

de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182(1), 375–385. <https://doi.org/10.1534/genetics.109.101501>

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407–499. <https://doi.org/doi:10.1214/009053604000000067>

Euler, K. N., Hauck, S. M., Ueffing, M., & Deeg, C. A. (2013). Bovine neonatal pancytopenia - Comparative proteomic characterization of two BVD vaccines and the producer cell surface proteome (MDBK). *BMC Veterinary Research*, 9. <https://doi.org/10.1186/1746-6148-9-18>

Fernando, R. L., & Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, 21(4), 467. <https://doi.org/10.1186/1297-9686-21-4-467>

Fiehn, O., Timothy Garvey, W., Newman, J. W., Lok, K. H., Hoppel, C. L., & Adams, S. H. (2010). Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS ONE*, 5(12), 1–10. <https://doi.org/10.1371/journal.pone.0015234>

Frigolet, M. E., & Gutiérrez-Aguilar, R. (2017). Ciencias “ómicas”, ¿cómo ayudan a las ciencias de la salud? *Revista Digital Universitaria*, 18(7), 0–15. <https://doi.org/10.22201/codeic.16076079e.2017.v18n7.a3>

- Fundación Instituto Roche. (n.d.). Glosario de genética. Retrieved December 7, 2020, from <https://www.instituto Roche.es/recursos/glosario/>
- Fundación Instituto Roche. (2019). *Ciencias ómicas*. Retrieved from https://www.instituto Roche.es/static/archivos/Informes_anticipando_CIENCIAS_OMICAS.pdf
- Gianola, D., & Rosa, G. J. M. (2015). One hundred years of statistical developments in animal breeding. *Annual Review of Animal Biosciences*, 3, 19–56. <https://doi.org/10.1146/annurev-animal-022114-110733>
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6), 323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Hao, D., Wang, X., Wang, X., Thomsen, B., Kadarmideen, H. N., Lan, X., ... Chen, H. (2020). Transcriptomic changes in bovine skeletal muscle cells after resveratrol treatment. *Gene*, 754(November 2019), 144849. <https://doi.org/10.1016/j.gene.2020.144849>
- Hao, Q., Zhu, Z., Xu, D., Liu, W., Lyu, L., & Li, P. (2019). Proteomic characterization of bovine granulosa cells in dominant and subordinate follicles. *Hereditas*, 156, 21. <https://doi.org/10.1186/s41065-019-0097-5>
- Hayes, B. J., Chamberlain, A. J., McPartlan, H., Macleod, I., Sethuraman, L., & Goddard, M. E. (2007). Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genetical Research*, 89(04). <https://doi.org/10.1017/S0016672307008865>
- Hazel, L. N., & Lush, J. A. Y. L. (1942). The efficiency of three methods of selection. *Journal of Heredity*, 33(11), 7393–399. <https://doi.org/10.1093/oxfordjournals.jhered.a105102>
- Henderson, C. R. (1953). Estimation of variance and covariance components.

Biometrics, 9(2), 226. <https://doi.org/10.2307/3001853>

Henderson, C. R. (1963). Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, 892, 141–163. Retrieved from <http://morotalab.org/literature/pdf/henderson1963.pdf>

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2), 423. <https://doi.org/10.2307/2529430>

Henderson, C. R. (1984). Applications of linear models in animal breeding. In *University of Guelph*. Retrieved from chrome-extension://efaidnbmnnnibpcajpcgiclfendmkaj/viewer.html?pdfurl=https%3A%2F%2Fwww.ancp.org.br%2Fwp-content%2Fuploads%2F2011%2F02%2FLivro-Applications-of-Linear-Models-in-Animal-Breeding.pdf&clen=2469051&chunk=true

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>

Huang, Y., Shen, L., Jiang, J., Xu, Q., Luo, Z., Luo, Q., ... Cao, S. (2019). Metabolomic profiles of bovine mammary epithelial cells stimulated by lipopolysaccharide. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-55556-2>

Infantes-Lorenzo, J. A., Moreno, I., Rialde, M. D. L. Á., Roy, Á., Villar, M., Romero, B., ... Domínguez, M. (2017). Proteomic characterisation of bovine and avian purified protein derivatives and identification of specific antigens for serodiagnosis of bovine tuberculosis. *Clinical Proteomics*, 14(1), 1–10. <https://doi.org/10.1186/s12014-017-9171-z>

Iowa State University. (n.d.). Robert Bakewell | Department of Animal Science. Retrieved December 12, 2020, from

<https://www.ans.iastate.edu/about/history/people/robert-bakewell>

- Jacquín, L., Cao, T. V., & Ahmadi, N. (2016). A unified and comprehensible view of parametric and kernel methods for genomic prediction with application to rice. *Frontiers in Genetics*, 7(AUG), 1–16. <https://doi.org/10.3389/fgene.2016.00145>
- Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9(2), 166–177. <https://doi.org/10.1093/bfpg/elq001>
- Kell, D. B., & Oliver, S. G. (2016). The metabolome 18 years on: a concept comes of age. *Metabolomics*, 12(10), 1–8. <https://doi.org/10.1007/s11306-016-1108-4>
- Kinkead, R. A., Elliott, C. T., Cannizzo, F. T., Biolatti, B., Gadaj, A., & Mooney, M. H. (2018). Plasma metabolomic profiling based detection of drug specific responses to different bovine growth promoting regimes. *Food Control*, 86, 146–158. <https://doi.org/10.1016/j.foodcont.2017.10.036>
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. <https://doi.org/10.2307/1913643>
- Kouassi, J. J., Dervilly-Pinel, G., Chéreau, S., Biancotto, G., Monteau, F., Elliott, C. T., & Le Bizec, B. (2015). First insights into serum metabolomics of trenbolone/estradiol implanted bovines; screening model to predict hormone-treated and control animals' status. *Metabolomics*, 11(5), 1184–1196. <https://doi.org/10.1007/s11306-015-0775-x>
- Kropp, J., Carrillo, J. A., Namous, H., Daniels, A., Salih, S. M., Song, J., & Khatib, H. (2017). Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. *BMC Genomics*, 18(1), 1–15. <https://doi.org/10.1186/s12864-017-3673-y>
- Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Appl. Statist*, 41(1), 191–201. Retrieved from

http://gs3sr3zm5k.search.serialssolutions.com.ezproxy.conricyt.org/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rft_id=info%3Aid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=RIDGE+

Le, T. T., Deeth, H. C., & Larsen, L. B. (2017). Proteomics of major bovine milk proteins: Novel insights. *International Dairy Journal*, 67, 2–15. <https://doi.org/10.1016/j.idairyj.2016.11.016>

Lee, A. H., & Silvapulle, M. J. (1988). Ridge estimation in logistic regression. *Communications in Statistics -Simulation and Computation*, 17(4), 1231–1257. <https://doi.org/10.1080/03610918808812723>

Lippolis, J. D., Powell, E. J., Reinhardt, T. A., Thacker, T. C., & Casas, E. (2019). Symposium review: Omics in dairy and animal science-Promise, potential, and pitfalls. *Journal of Dairy Science*, 102(5), 4741–4754. <https://doi.org/10.3168/jds.2018-15267>

López Cruz, M. A. (2012). *Aplicación del Elastic Net LASSO y modelos relacionados en Selección Genómica basados en Marcadores Moleculares*. (Colegio de Postgraduados (COLPOS)). Colegio de Postgraduados (COLPOS). Retrieved from http://www.biblio.colpos.mx:8080/jspui/bitstream/handle/10521/256/Sanchez_Borja_M_DC_Fitosanidad_2010.pdf?sequence=1

Ma, D., Yu, Q., Hedrick, V. E., Cooper, B. R., Paschoal Sobreira, T. J., Oh, J. H., ... Kim, Y. H. B. (2020). Proteomic and metabolomic profiling reveals the involvement of apoptosis in meat quality characteristics of ovine *M. longissimus* from different callipyge genotypes. *Meat Science*, 166(January), 108140. <https://doi.org/10.1016/j.meatsci.2020.108140>

Martínez, C. A., Manrique, C., & Elzo, M. A. (2012). Cattle genetic evaluation: a historical perception. *Revista Colombiana de Ciencias Pecuarias*, 25, 293–311.

Retrieved from <https://www.redalyc.org/pdf/2950/295023555014.pdf>

- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense markers maps. *Genetics*. <https://doi.org/11290733>
- Moore, S. C., Matthews, C. E., Sampson, J. N., Stolzenberg-Solomon, R. Z., Zheng, W., Cai, Q., ... Cross, A. J. (2014). Human metabolic correlates of body mass index. *Metabolomics*, 10(2), 259–269. <https://doi.org/10.1007/s11306-013-0574-1>
- Moran-Ramos, S., Ocampo-Medina, E., Gutierrez-Aguilar, R., Maclás-Kauffer, L., Villamil-Ramírez, H., López-Contreras, B. E., ... Canizales-Quinteros, S. (2017). An amino acid signature associated with obesity predicts 2-year risk of hypertriglyceridemia in school-age children. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-05765-4>
- Nascimento, M., e Silva, F. F., de Resende, M. D. V., Cruz, C. D., Nascimento, A. C. C., Viana, J. M. S., ... Barroso, L. M. A. (2017). Regularized quantile regression applied to genome-enabled prediction of quantitative traits. *Genetics and Molecular Research*, 16(1). <https://doi.org/10.4238/GMR16019538>
- Nicolás P., E. (2013). *Método De Regresión Bayesiana Para Selección Genómica*. (Colegio de Postgraduados (COLPOS)). Colegio de Postgraduados (COLPOS). Retrieved from http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2089/Nicolas_Popoca_E_MC_Estadistica_2013.pdf?sequence=1
- Park, S., Sadanala, K. C., & Kim, E. K. (2014). A metabolomic approach to understanding the metabolic link between obesity and diabetes. *Molecules and Cells*, 38(7), 587–596. <https://doi.org/10.14348/molcells.2015.0126>
- Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American*

Statistical Association, 103(482), 681–686.
<https://doi.org/10.1198/016214508000000337>

- Peddinti, D., Nanduri, B., Kaya, A., Feugang, J. M., Burgess, S. C., & Memili, E. (2008). Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Systems Biology*, 2, 1–13. <https://doi.org/10.1186/1752-0509-2-19>
- Pérez-Rodríguez, P., Montesinos-López, O. A., Montesinos-López, A., & Crossa, J. (2020). Bayesian regularized quantile regression: A robust alternative for genome-based prediction of skewed data. *Crop Journal*. <https://doi.org/10.1016/j.cj.2020.04.009>
- Picardi, L. A. (2017). Los pasos de la genética animal en el siglo XXI. *Ciencia e Investigación*, 67, 43–47. Retrieved from <http://aargentinapciencias.org/wp-content/uploads/2018/01/RevistasCel/tomo67-1/4-Picardi-cei67-1-3.pdf>
- Rajacic, P., Weersink, A., Navabi, A., & Peter Pauls, K. (2016). Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. *Euphytica*, 210(2), 259–276. <https://doi.org/10.1007/s10681-016-1716-0>
- Schaefer, R. L., Roi, L. D., & Wolfe, R. A. (1984). A Ridge logistic estimator. *Commun. Statist.-Theor. Meth.*, 13(1), 99–113. <https://doi.org/10.1080/03610928408828664>
- Schaeffer, L. R. (2006). Strategy for applying genome wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123(4), 218–223.
- Schroyen, M., & Tuggle, C. K. (2015). Current transcriptomics in pig immunity research. *Mammalian Genome*, 26(1-2), 1-20. <https://doi.org/10.1007/s00335-014-9549-4>
- Stella, R., Dervilly-Pinel, G., Bovo, D., Mastroianni, E., Royer, A. L., Angeletti, R., ... Biancotto, G. (2017). Metabolomics analysis of liver reveals profile disruption in

- bovines upon steroid treatment. *Metabolomics*, 13(7), 1–13. <https://doi.org/10.1007/s11306-017-1220-0>
- Tibshirani, R. (1996). Regression selection and shrinkage via the Lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288. <https://doi.org/10.2307/2346178>
- Tirados F., S. P. (2001). La mejora genética animal en la segunda mitad del siglo xx. *Archivos de Zootecnia*, 50(192), 517–546. Retrieved from <https://www.redalyc.org/articulo.oa?id=49519206>
- Uhde, K., Van Tol, H. T. A., Stout, T. A. E., & Roelen, B. A. J. (2018). Exposure to elevated glucose concentrations alters the metabolomic profile of bovine blastocysts. *PLoS ONE*, 13(6). <https://doi.org/10.1371/journal.pone.0199310>
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Richard, J., Sutton, G. G., ... France, V. Di. (2001). The sequence of the Hum. *Science*, 291, 1304–1350. Retrieved from http://www.jstor.org.conricyt.remotexs.co/stable/3083494?pq-origsite=summon&seq=1#metadata_info_tab_contents
- Wang, X., Yu, Z., Zhao, X., Han, R., Huang, D., Yang, Y., & Cheng, G. (2020). Comparative proteomic characterization of bovine milk containing β -casein variants A1A1 and A2A2, and their heterozygote A1A2. *Journal of the Science of Food and Agriculture*, (July). <https://doi.org/10.1002/jsfa.10684>
- Wiltafsky, M. K., Pfaffl, M. W., & Roth, F. X. (2010). The effects of branched-chain amino acid interactions on growth performance, blood metabolites, enzyme kinetics and transcriptomics in weaned pigs. *British Journal of Nutrition*, 103(7), 964–976. <https://doi.org/10.1017/S0007114509992212>
- Yu, Q., Tian, X., Shao, L., Li, X., & Dai, R. (2019). Targeted metabolomics to reveal muscle-specific energy metabolism between bovine longissimus lumborum and psoas major during early postmortem periods. *Meat Science*, 156(May), 166–173. <https://doi.org/10.1016/j.meatsci.2019.05.029>

- Yudin, N. S., Lukyanov, K. I., Voevoda, M. I., & Kolchanov, N. A. (2016). Application of reproductive technologies to improve dairy cattle genomic selection. *Russian Journal of Genetics: Applied Research*, 6(3), 321–329. <https://doi.org/10.1134/S207905971603014X>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

3. SINGLE-STEP GENOMIC EVALUATION FOR GROWTH TRAITS IN A MEXICAN BRAUNVIEH CATTLE POPULATION

Jonathan Emanuel Valerio-Hernández¹, Agustín Ruíz-Flores¹, Mohammad Ali Nilforooshan³, Paulino Pérez-Rodríguez²

¹ Departamento de Zootecnia, Universidad Autónoma Chapingo, Km 38.5 Carretera México-Texcoco. Chapingo, Estado de México. CP 56227

² Programa de Estadística, Colegio de Postgraduados. Montecillo, Estado de México. CP 56264

³ Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand

Submitted to *Animal Bioscience*

Title: Single-step genomic evaluation for growth traits in a Mexican Braunvieh cattle population

Type of Manuscript: Article

Abstract

Objective: The objective was to compare (pedigree-based) BLUP, genomic BLUP (GBLUP), and single-step GBLUP (ssGBLUP) methods for genomic evaluation of growth traits in a Mexican Braunvieh cattle population. **Methods:** Birth (BW), weaning (WW), and yearling weight (YW) data of a Mexican Braunvieh cattle population were analyzed with BLUP, GBLUP, and ssGBLUP methods. These methods are differentiated by the additive genetic relationship matrix included in the model and the animals under evaluation. The predictive ability of the model was evaluated using random partitions of the data in training and testing sets, consistently predicting about 20% of genotyped animals on all occasions. For each partition, the Pearson correlation coefficient between adjusted phenotypes for fixed effects and non-genetic random effects and the estimated breeding values (EBV) were computed. **Results:** The random contemporary group (CG) effect explained about 50, 45, and 35% of the phenotypic variance in BW, WW, and YW, respectively. For the three methods, the CG effect explained the highest proportion of the phenotypic variances (except for YW-GBLUP). The heritability estimate obtained with GBLUP was the lowest for BW, while the highest heritability was obtained with BLUP. For WW, the highest heritability estimate was obtained with BLUP, the estimates obtained with GBLUP and ssGBLUP were similar. For YW, the heritability estimates obtained with GBLUP and BLUP were similar, and the lowest heritability was obtained with ssGBLUP. Pearson correlation coefficients between adjusted phenotypes for non-genetic effects and EBVs were the highest for BLUP, followed by ssBLUP and GBLUP. **Conclusion:** The successful implementation of genetic evaluations that include genotyped and non-genotyped animals in our study indicate a promising method for use in genetic improvement programs of Braunvieh cattle. Our findings showed that simultaneous evaluation of genotyped and non-genotyped animals improved prediction accuracy for growth traits even with a limited number of genotyped animals.

Editorial members
Animal Bioscience Editorial Office
Room 708 Sammo Sporex, 23, Sillim-ro 59-gil,
Gwanak-gu, Seoul 08776, Korea

Tel : +82-2-888-6558
Fax : +82-2-888-6559
E-mail : animbiosci@gmail.com
Website: <https://submit.animbiosci.org/>

Room 708 Sammo Sporex, 23, Sillim-ro 59-gil, Gwanak-gu, Seoul 08776, Korea

Copyright© Animal Bioscience.

1 **Single-step genomic evaluation for growth traits in a Mexican Braunvieh**
2 **cattle population**

3
4 Jonathan Emanuel *Valerio-Hernández*¹, Agustín *Ruíz-Flores*^{1,a}, Mohammad Ali *Nilforooshan*³,
5 Paulino *Pérez-Rodríguez*^{2,a}

6
7 ^a **Corresponding authors: Paulino Pérez Rodríguez**

8 Tel: +52-595-952-0200, Ext. 1459, Fax: +52-595-952-0200, Email: perpdgo@colpos.mx

9 **Agustín Ruíz-Flores**

10 Tel: +52-595-952-1500, Ext. 1683, Fax: +52-595-952-1621, Email: aruizf@chapingo.mx

11
12 ¹ Departamento de Zootecnia, Universidad Autónoma Chapingo, Km 38.5 Carretera México-
13 Texcoco. Chapingo, Estado de México. CP 56227

14 ² Programa de Estadística, Colegio de Postgraduados. Montecillo, Estado de México. CP 56264

15 ³ Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand

16 **ORCID**

17 Jonathan E. Valerio-Hernández 0000-0001-5043-8914

18 Paulino Pérez-Rodríguez 0000-0002-3202-1784

19 Mohammad Ali Nilforooshan 0000-0003-0339-5442

20 Agustín Ruíz-Flores 0000-0001-8267-2107

Single-step genomic evaluation for growth traits in a Mexican Braunvieh cattle population

ABSTRACT

Objective: The objective was to compare (pedigree-based) BLUP, genomic BLUP (GBLUP), and single-step GBLUP (ssGBLUP) methods for genomic evaluation of growth traits in a Mexican Braunvieh cattle population.

Methods: Birth (BW), weaning (WW), and yearling weight (YW) data of a Mexican Braunvieh cattle population were analyzed with BLUP, GBLUP, and ssGBLUP methods. These methods are differentiated by the additive genetic relationship matrix included in the model and the animals under evaluation. The predictive ability of the model was evaluated using random partitions of the data in training and testing sets, consistently predicting about 20% of genotyped animals on all occasions. For each partition, the Pearson correlation coefficient between adjusted phenotypes for fixed effects and non-genetic random effects and the estimated breeding values (EBV) were computed.

Results: The random contemporary group (CG) effect explained about 50, 45, and 35% of the phenotypic variance in BW, WW, and YW, respectively. For the three methods, the CG effect explained the highest proportion of the phenotypic variances (except for YW-GBLUP). The heritability estimate obtained with GBLUP was the lowest for BW, while the highest heritability was obtained with BLUP. For WW, the highest heritability estimate was obtained with BLUP, the estimates obtained with GBLUP and ssGBLUP were similar. For YW, the heritability estimates obtained with GBLUP and BLUP were similar, and the lowest heritability was obtained with ssGBLUP. Pearson correlation coefficients between adjusted phenotypes for non-genetic effects and EBVs were the highest for BLUP, followed by ssBLUP and GBLUP.

Conclusion: The successful implementation of genetic evaluations that include genotyped and non-genotyped animals in our study indicate a promising method for use in genetic improvement programs of Braunvieh cattle. Our findings showed that simultaneous evaluation of genotyped and non-genotyped animals improved prediction accuracy for growth traits even with a limited number of genotyped animals.

Keywords: BLUP, Braunvieh, Genomic selection, GBLUP, ssGBLUP.

INTRODUCTION

Traditionally, genetic merits of individuals have been predicted (EBV) using pedigree information [1]. Genomic information can be used to predict EBV. When this occurs, using GBLUP [2], it is called genomically-enhanced EBV or GEBV. Genomic selection based on single nucleotide polymorphisms (SNP) accelerates genetic progress by increasing the accuracy of predictions, reducing generation interval, and controlling inbreeding [3].

Nowadays, genomic evaluations are widely used in domestic animal species thanks to the technological advances and a reduction in genotyping costs. Availability of dense SNP panels has made genome-wide or genomic selection evaluation methods effective. Most of the methods for genomic evaluations are usually based on variants of genome-wide association studies for estimating the effects of markers or haplotypes. Differences among the approaches reside in the *a priori* distribution assumed for markers or haplotype effects [4]. However, training these genomic evaluation models generally requires the entire population to be both phenotyped and genotyped, which is not always true or feasible [5]. A more recent method [2] evaluates a population of individuals, not all of which are genotyped. This genomic evaluation method called the single-step genomic BLUP (ssGLUP) was developed by Christensen and Lund [6] and Aguilar et al. [7]. The ssGBLUP method allows the inclusion of both genotyped and non-genotyped animals and uses pedigree and genotype information simultaneously. The two research teams [6,7], using different analytical approaches, arrived at the same formulation for ssGBLUP. The central idea of the method is to create an improved additive genetic relationship matrix (**H**), from the pedigree-based (**A**) and the marker-based (**G**) additive genetic relationship matrices. An advantage of ssGBLUP is that it is extendable to all forms of BLUP by simply replacing the $\mathbf{A}^{-1}$ matrix in BLUP [1] with the $\mathbf{H}^{-1}$ matrix, and EBVs are obtained for both genotyped and non-genotyped animals. A statistically equivalent model to ssGBLUP, which we refer to as ssSNPBLUP, was developed by Liu et al. [8]. The model directly estimates marker effects, which avoids creating and inverting the genomic relationship matrix. Inverting the genomic relationship matrix can be challenging for a large number of genotyped animals.

Liu et al. [8] mentioned that the proposed method works well with populations of any size. Other methods equivalent to ssSNPBLUP were developed by Fernando et al. [9,10].

The ssGBLUP method has been widely used in both plant and animal breeding [11,12,13]. Increases in accuracy of predictions by ssGBLUP have been reported, relative to those obtained with BLUP and GBLUP [5,12,14]. The ssGBLUP method has been used in Holstein dairy cattle [15,16], Nordic Red dairy cattle [17], dual-purpose cattle [18], swine [12], turkeys [19], sheep [20,21], and goats [22]. However, to our knowledge the ssGBLUP method has not yet been used in the genetic evaluation of Braunvieh cattle.

In Mexico, BLUP phenotype-based genetic evaluations of Braunvieh cattle are performed on a regular basis [23]. The objective of this study was to compare ssGBLUP, BLUP, and GBLUP, using phenotypic records for growth traits of a Mexican Braunvieh cattle population.

MATERIAL AND METHODS

Animals

Phenotypic and pedigree information of the Braunvieh cattle population came from the database of the Mexican Association of Braunvieh Purebreeders (Asociación Mexicana de Criadores de Ganado Suizo de Registro, Mexico City). Birth weight (BW), weaning (at 218 days of age) weight (WW), and yearling weight (YW) phenotypes were used. Data on 10 generations back were extracted. Animals were born from 1998 to 2016.

For the genomic information, hair samples were collected from 300 animals in eight herds located in Eastern, Central, and Western Mexico. The samples were genotyped at GeneSeek (Lincoln, <https://www.neogen.com/>, NE, USA), using the GeneSeek® Genomic Profiler Bovine LDv.4 panel, with 30,000 and 50,000 SNP markers, 150 animals with each Chip.

Phenotypes

Data editing for BW started with 32,159 records. Outliers were removed (animals with phenotypes higher or lower than the average phenotype in the database $\pm$ 3 standard deviations). Records from animals without information of herd or dam age were deleted as well. Animals not genetically related

to the 300 genotyped animals were discarded because the solutions for non-genotyped animals not related to genotyped animals would remain the same between BLUP and ssGBLUP. Contemporary groups (CG) were defined by combining the effects of herd (8 herds), year (from 1998 to 2016), and season of birth. Birth seasons were defined considering the Julian calendar, from 80 to 171 d, spring; from 172 to 264 d, summer; from 265 to 354 d, autumn; from 355 to 366 d, and from 1 to 79 d, winter. The days were counted continuously from the beginning of the Julian period, starting from the first of January each year. Records of animals in CG with fewer than two individuals or with phenotypic variance equal to zero were discarded. Data edition resulted in 330 BW records for further analysis. Of the phenotyped animals, 232 had been genotyped.

For WW data, editing started with 29,142 records. Those records with erroneous age at weaning, were outside the range average ± 3 standard deviations, or had no information on management, herd, or age of dam were discarded. Animals not genetically related to the 300 genotyped animals were deleted as well. Contemporary groups for WW were defined in the same way as for BW with 6 herds and birth years from 1998 to 2016. Three management groups were defined: calves fed milk from their dam, calves fed milk from their dam plus balanced feed, and calves fed milk from their dam and a wet nurse plus balanced feed. Records of animals in CG with fewer than two animals and variance equal to zero were discarded as well. At the end of data editing, 267 WW records were left for further analysis. Of the phenotyped animals, 218 had been genotyped.

Data editing for YW started with 19,971 records. Those with discordant age, or outside the average ± 3 standard deviations, with no herd or management information were deleted. Animals not genetically related to the 300 genotyped animals were discarded as well. Contemporary groups were defined in the same way as for WW. Three management groups for YW were defined: animals kept under grazing, animals that grazed and received balanced feed, and confined animals that received balanced feed. Records of animals in CG with fewer than two individuals and variance equal to zero were deleted as well. At the end of data editing, 232 records for YW were available for subsequent analyses. Of the phenotyped animals, 191 had been genotyped.

Genotypes

Data from 300 genotyped animals (236 females and 64 males) born between 2001 and 2016 were used. The SNPs in common between the 30K and 50K Chips were used (12,835 SNPs). Proportions of missing values for each marker and for each individual were calculated. The average of missing values by individual was 2.09% with a standard deviation of 7.50%. The average call rate (non-missing proportion for each marker) was 97.90% with a standard deviation of 4.66%. Markers with a call rate below 95% were removed. Missing marker genotypes were imputed at random, generating samples of $Binomial(2, \hat{p})$ distribution with $\hat{p}$ computed from observed marker genotypes. Monomorphic markers or those with minor allele frequency less than 0.04 were removed. After all the cleaning and quality control process, 11,648 of the 12,835 SNPs in common between the two Chips were available for further analysis.

Calculation of the genetic relationship matrices

The inverse of the additive genetic relationship matrix ($\mathbf{A}^{-1}$) was obtained from the pedigree information, using the function 'getAInv' from R package 'pedigreemm' [24]. The genomic relationship matrix was calculated as described by López-Cruz et al. [25] and Pérez-Rodríguez et al. [13]. Briefly, $\mathbf{G} = \mathbf{W}\mathbf{W}'/n$, where $\mathbf{W}$ is the matrix of recoded marker genotypes centered and standardized by column, and n is the number of markers. Because all markers were standardized to a variance of one, the average of the diagonal elements of $\mathbf{G}$ was about one, so that both matrices $\mathbf{A}$ and $\mathbf{G}$ have similar scaling.

The inverse of the combined pedigree and genomic additive relationship matrix, including genotyped and non-genotyped animals ($\mathbf{H}^{-1}$), is obtained directly without the need to form $\mathbf{H}$. This matrix is defined as [7,12]:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_a^{-1} - \mathbf{A}_{gg}^{-1} \end{bmatrix},$$

where, $\mathbf{A}_{gg}$ is the submatrix of $\mathbf{A}$ for genotyped animals, $\mathbf{G}_a = \beta\mathbf{G} + \alpha$; β and α were obtained by solving the equations:

$$\begin{cases} Avg(diag(\mathbf{G}))\beta + \alpha = Avg(diag(\mathbf{A}_{gg})) \\ Avg(\mathbf{G})\beta + \alpha = Avg(\mathbf{A}_{gg}) \end{cases}.$$

The obtained α (0.0296) and β (0.9809) values set the mean of $\mathbf{G}_a$ to the mean of $\mathbf{A}_{gg}$, and the mean of $diag(\mathbf{G}_a)$ to the mean of $diag(\mathbf{A}_{gg})$. Figure 1 shows the distribution of the diagonal elements for matrices $\mathbf{G}$, $\mathbf{A}$ and $\mathbf{H}$.

Statistical analysis

Growth traits BW, WW, and YW were analyzed with different univariate models and three alternative methods, BLUP, GBLUP, and ssGBLUP. The most important difference among these methods is the inverse relationship matrices used ($\mathbf{A}^{-1}$, $\mathbf{G}^{-1}$, $\mathbf{H}^{-1}$) and the animals under evaluation (only genotyped animals for GBLUP).

Analyses of the three traits were performed with the linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{c} + \mathbf{Z}_2\mathbf{a} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is the vector of phenotypes, $\mathbf{X}$ is the incidence matrix relating phenotypes to fixed effects (3, 4, and 4 fixed effects for BW, WW, and YW), $\mathbf{Z}_1$ is the incidence matrix relating phenotypes to random contemporary groups (54, 43, and 37 CG for BW, WW, and YW in GBLUP, and 96, 66, and 59 CG for BW, WW, and YW in BLUP and ssGBLUP), $\mathbf{Z}_2$ is the incidence matrix relating phenotypes to random additive genetic effects, and $\mathbf{e}$ is the vector of random residual effects. Additive genetic, CG, and residual variances were $Var(\mathbf{a}) = \Delta\sigma_a^2$, $Var(\mathbf{c}) = \mathbf{I}\sigma_{cg}^2$, and $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$, respectively. Matrix Δ is $\mathbf{H}$, $\mathbf{A}$, or $\mathbf{G}$, depending on the method of analysis, and σ_a^2 , σ_{cg}^2 , and σ_e^2 are the corresponding variances. Variance components for each method were estimated via the R package ‘BGLR’ [26] using the Bayesian framework. Hyper-parameters were set automatically according to internal rules [26]. Inferences were based on 25,000 MCMC iterations obtained after discarding 25,000 samples taken as burn-in and with a thin set to 10. The fixed effects were age of dam and sex for BW; age of dam, sex, weaning age, and management group for WW; and age of dam, sex, yearling age, and management group for YW.

Validation analysis

GBLUP model

To study the predictive ability of the models, we partitioned the data at random to 80% of the observations as the training set and the remaining 20% as the testing set. We generated 100 partitions of the data at random and fitted model (1) using the BGLR package. The same parameters and hyperparameters described in the previous section were used for the variance components estimation. For each partition, we computed the Pearson correlation coefficient between the phenotypes adjusted for fixed effects and non-genetic random effects and EBV. The average of the Pearson correlation coefficients and their standard deviations were calculated. For each of the 100 random validation partitions, we kept a record of which individuals were assigned to the testing set to validate the same set of individuals with BLUP and ssGBLUP methods.

BLUP and ssGBLUP methods

To study the predictive ability of these methods, the data was partitioned into training and testing subsets. We assigned the same individuals predicted with the GBLUP model, as well as the rest of the individuals, to the testing set to have a common set of individuals whose breeding values needed to be predicted. We fitted the models and calculated Pearson correlation coefficients between the adjusted phenotypes and the EBVs using the same strategy as that used in the GBLUP model.

RESULTS

Table 1 shows the estimates of additive genetic (σ_a^2), contemporary group (σ_{cg}^2) and residual variances for BW, WW, and YW. The random contemporary group effect explained about 50, 45, and 35% of the phenotypic variance for BW, WW, and YW, respectively. For the three traits and the three methods, σ_{cg}^2 explained the highest proportion of the phenotypic variance, except for YW-GBLUP. Heritability estimates (h^2) are also presented in Table 1. For BW, the heritability estimate obtained with GBLUP was the lowest, while the highest was obtained with BLUP. For WW, the highest heritability estimate was obtained with BLUP; the estimates obtained with GBLUP and ssGBLUP

were similar. For YW, the heritability estimates obtained with GBLUP and BLUP were similar; the lowest was obtained with ssGBLUP. Within traits, the heritability estimates obtained with the three alternative methods were within a narrow range, except for BW-GBLUP ($h^2 = 0.132$). The high standard errors associated with the variance components (Table 1) might be due to the small sample size.

Table 2 shows the Pearson correlation coefficients between the adjusted phenotypes and the EBVs. For BW, the lowest and the highest correlations were obtained with GBLUP and BLUP, respectively. BLUP yielded consistently higher estimates than either GBLUP or ssGBLUP. Figure 2 shows graphically the results of the validation; each bar represents a correlation estimate and the error bars represent the associated standard errors computed by dividing the standard deviation in parenthesis in Table 2 by 10, the square root of the number of random partitions evaluated.

DISCUSSION

A potential benefit of ssGBLUP over BLUP and GBLUP is that it utilizes data used in both BLUP and GBLUP, simultaneously. Thus, ssGBLUP and BLUP use phenotypes and pedigrees from genotyped and non-genotyped animals. As GBLUP is limited to genotyped animals, only phenotypes of genotyped animals were used.

The statistical models for BW and WW typically include a maternal genetic effect. However, we did not have enough information to estimate this effect to include it in the model due to the small sample size. Consequently, this will lead to an increase in the uncertainty of predictions.

For the three traits and the three methods tested, σ_{cg}^2 had the greatest magnitude among the estimated variance components. Among all the trait-method combinations, σ_a^2 was always the greatest with BLUP (Table 1). In contrast, σ_a^2 was consistently the lowest for GBLUP. Estimation of the variance components and genetic parameters is useful for planning selection in breeding programs; additive genetic variance has long been regarded as the most important.

The heritability estimates were within the range of what has been reported in other studies. For example, heritability estimates for several cattle breeds using BLUP (A^{-1}) were in ranges from 0.15 to 0.36, from 0.10 to 0.27, and from 0.18 to 0.30 for BW, WW and YW, respectively [27,28]. Pearson correlation estimates between the adjusted phenotypes and the EBVs were highest for BLUP, followed by ssGBLUP. The estimates obtained with GBLUP and ssGBUP were quite similar for YW. The reported Pearson correlation coefficients between corrected phenotypes and EBVs allow us to determine how well a model predicts the additive genetic merits. In this study, BLUP gave the best predictions. We also calculated Pearson correlation coefficients between observed and predicted phenotypes. This correlation determines the ability of the model to predict the phenotype of interest, given the information provided in the model. Other studies [12,29,30] reported better model fits by including genomic information; however, genomic information in this study was limited. Table 2 presents the Pearson correlation coefficients between corrected phenotypes and EBVs together with the standard errors around the correlation coefficients. The number of individuals in the training set for methods BLUP and ssGBLUP was larger than in the training set for GBLUP, and the number of individuals predicted in the three cases were the same. The sample size in the training set could result in better predictions in the testing set when datasets are small. Other studies obtained higher prediction accuracies using ssGBLUP [11,12,14,19]. However, these authors used larger populations than that of our study. We used 11,648 SNPs to compute **G** and **H** matrices, while in the mentioned studies 53,455, 45,818, 25,720, and 57,000 SNP markers, respectively, were used. Even though, the number of SNPs was not as large as in those studies, the correlation estimates results (BLUP vs. ssGBLUP) suggest that the information provided by 11,648 markers is valuable. Christensen et al. [12] studying average daily gain and feed conversion in Danish Duroc swine, reported that ssGBLUP and GBLUP yielded more accurate predictions than BLUP. These advantages held true in both uni- and bivariate analyses. Additionally, they found that the ssGBLUP method resulted in more accurate predictions for non-genotyped animals [12]. These authors concluded that

ssGBLUP yields more accurate predictions for genotyped animals than BLUP, and similar accuracy compared with GBLUP for genotyped animals. Ma et al. [29] and Park et al. [30] reported slightly higher accuracies for ssGBLUP than for BLUP for five carcass traits of Hanwoo cattle.

CONCLUSION

The literature reports that ssGBLUP outperforms pedigree and marker-based methods. The gains in some cases are slightly higher and depend on several factors, for example, number of individuals genotyped and non-genotyped, sample size, number of markers, and heritability of the traits. In our case, the number of genotyped animals was small compared with other studies, and this is perhaps the main reason that we did not observe a clear advantage of ssGBLUP over BLUP. However, the results are encouraging, and we will continue genotyping and phenotyping more animals (especially those with more relatives in the population) in the years to come to increase our sample sizes. Another possibility for improving the benefits of ssGBLUP is making use of genotypes on a larger set of markers by genotyping more animals for both 30K and 50K panels to make imputation to the 50K panel possible.

CONFLICT OF INTEREST

The authors declare no conflict of interest with any financial organization regarding the matters discussed in the manuscript.

ACKNOWLEDGEMENTS

The authors acknowledge the Asociación Mexicana de Criadores de Ganado Suizo de Registro (AMCGSR, Mexico City) and the collaborating breeders for permitting the use of their databases for this study. Special thanks to the Consejo Nacional de Ciencia y Tecnología, Mexico, for providing financial support to the first author for his Doctorate studies.

REFERENCES

1. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 1975;31(2):423. <https://doi.org/10.2307/2529430>
2. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci* 2008;91:4414-4423. <https://doi.org/10.3168/jds.2007-0980>
3. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense markers maps. *Genetics* 2001;157(4):1819-29. <https://doi.org/10.1093/genetics/157.4.1819>
4. Gianola D. Priors in whole-genome regression: The bayesian alphabet. *Genetics* 2013;194(3):573-596. <https://doi.org/10.1534/genetics.113.151753>
5. Legarra A, Christensen OF, Aguilar I, Misztal I. Single Step, a general approach for genomic selection. *Livest Sci* 2014;166(1):54-65. <https://doi.org/10.1016/j.livsci.2014.04.029>
6. Christensen O, Lund M. Genomic relationship matrix when some animals are not genotyped. *Genet Sel Evol* 2010;42(2):1-8. <http://www.gsejournal.org/content/42/1/2>
7. Aguilar I, Misztal I, Legarra A, Tsuruta S. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J Anim Breed Genet* 2011;28:422-428. <https://doi.org/10.1111/j.1439-0388.2010.00912.x>
8. Liu Z, Goddard ME, Reinhardt F, Reents R. A single-step genomic model with direct estimation of marker effects. *J Dairy Sci* 2014;97(9):5833–5850. <https://doi.org/10.3168/jds.2014-7924>
9. Fernando RL, Dekkers JCM, Garrick DJ. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet Sel Evol* 2014;46(1):1-13. <https://doi.org/10.1186/1297-9686-46-50>
10. Fernando RL, Cheng H, Golden BL, Garrick DJ. (2016). Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet Sel Evol* 2016;48(1):1–8. <https://doi.org/10.1186/s12711-016-0273-2>
11. Ashraf B, Edriss V, Akdemir D, Autrique E, Bonnett D, Crossa J, Janss L, Singh RG, Jannink JL. Genomic prediction using phenotypes from pedigreed lines with no marker data. *Crop Sci* 2016;56(3), 957-964. <https://doi.org/10.2135/cropsci2015.02.0111>
12. Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. Single-step methods for genomic evaluation in pigs. *Animal* 2012;6(10):1565-1571. <https://doi.org/10.1017/s1751731112000742>
13. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, Autrique E, de los Campos G, Burgueño J, Dreisigacker S. Single-step genomic and pedigree Genotype × Environment interaction models for predicting wheat lines in international environments. *Plant Genome* 2017;10(2):0. <https://doi.org/10.3835/plantgenome2016.09.0089>
14. Yoshida GM, Carneiro R, Rodríguez FH, Lhorente JP. Genomics Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic

- necrosis virus in rainbow trout. *Genomics* 2019;111(2):127–132.
<https://doi.org/10.1016/j.ygeno.2018.01.008>
15. Cornelissen MAMC, Mullaart E, Van der Linde C, Mulder HA. Estimating variance components and breeding values for number of oocytes and number of embryos in dairy cattle using a single-step genomic evaluation. *J Dairy Sci* 2017;100(6):4698-4705. <https://doi.org/10.3168/jds.2016-12075>
 16. Tiezzil F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS ONE*, 2015;10(2):1-15. <https://doi.org/10.1371/journal.pone.0114919>
 17. Koivula M, Strandén I, Pösö J, Aamand GP, Mäntysaari EA. Single-step genomic evaluation using multitrait random regression model and test-day data. *J Dairy Sci* 2015;98(4):2775-2784. <https://doi.org/10.3168/jds.2014-8975>
 18. Croué I, Ducrocq V. Genomic and single-step evaluations of carcass traits of young bulls in dual-purpose cattle. *J Anim Breed Genet*, 2017;134(4), 300–307. <https://doi.org/10.1111/jbg.12261>
 19. Abdalla EEA, Schenkel FS, Begli HE, Willems OW, van As P, Vanderhout R, Wood BJ, Baes CF. Single-step methodology for genomic evaluation in turkeys (*Meleagris gallopavo*). *Front Genet* 2019;10:1-8. <https://doi.org/10.3389/fgene.2019.01248>
 20. Nilforooshan MA. Application of single-step GBLUP in New Zealand Romney sheep. *Anim Prod Sci* 2020;60(9):1136-1144. <https://doi.org/10.1071/AN19315>
 21. Nilforooshan MA, Lee M. The quality of the algorithm for proven and young with various sets of core animals in a multibreed sheep population. *J Anim Sci* 2019;97(3):1090-1100. <https://doi.org/10.1093/jas/skz010>
 22. Carrilier C, Larroque H, Robert-Granié C. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. *Gen Sel Evol* 2014;46:67. <https://doi.org/10.1186/s12711-014-0067-3>.
 23. Núñez-Domínguez R, Ramírez-Valverde R, Ruíz-Flores A, Domínguez-Viveros J. Evaluación genética se sementales Suizo Europeo. Universidad Autónoma Chapingo, Departamento de Zootecnia, Chapingo, México. 2014.
 24. Vazquez AI, Bates DM, Rosa GJM, Gianola D, Weigel KA. Technical Note: An R package for fitting generalized linear mixed models in animal breeding. *J Anim Sci* 2010;88:497-504. <https://doi.org/10.2527/jas.2009-1952>
 25. López-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink J-L, Siingh R, Auatrique E, de los Campos, G. Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker ×

- Environment Interaction Genomic Selection Model. *G3* (Bethesda) 2015;5(4):569-582.
<https://doi.org/10.1534/g3.114.016097>
26. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 2014;198(2):483-495. <https://doi.org/10.1534/genetics.114.164442>
27. Montes-V D, Barragán-H W, Vergara-G O. Parametros genéticos de características productivas y reproductivas para ganado tipo carne en Colombia. *Revista Colombiana de Ciencia Animal-RECIA*, 2009;(2):302-318. <https://doi.org/10.24188/recia.v1.n2.2009.374>
28. Utrera ÁR, Velázquez GM, Murillo VEV, Bermúdez MM. Efectos genéticos para características de crecimiento de bovinos charolais y charbray Mexicanos estimados con modelos alternativos. *Rev Mex Cienc Pec* 2012;3(3):275-290.
29. Ma D, Yu Q, Hedrick VE, Cooper BR, Paschoal Sobreira TJ, Oh JH, Chun H, Kim YHB. Proteomic and metabolomic profiling reveals the involvement of apoptosis in meat quality characteristics of ovine *M. longissimus* from different callipyge genotypes. *Meat Sci* 2020;166:108140. <https://doi.org/10.1016/j.meatsci.2020.108140>
30. Park MN, Alam M, Kim S, Park B, Lee SH, Lee SS. Genomic selection through single-step genomic best linear unbiased prediction improves the accuracy of evaluation in Hanwoo cattle. *Asian-Australas J Anim Sci* 2020;33(10):1544-1557. <https://doi.org/10.5713/ajas.18.0936>

Table 1

Estimates of additive genetic variance (σ_a^2), contemporary group variance (σ_{cg}^2), residual variance (σ_e^2) and heritability (h^2) for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a Mexican Braunvieh cattle population, using best linear unbiased predictor (BLUP), genomic BLUP (GBLUP) and single-step GBLUP (ssGBLUP) methods. Standard errors for the variance parameters are shown in parenthesis.

		GBLUP		BLUP		ssGBLUP	
Component		Value	%	Value	%	Value	%
BW	σ_a^2	1.825	13.27	3.323	26.02	3.206	24.52
		(0.574)		(0.859)		(0.855)	
	σ_{cg}^2	8.845	64.32	6.348	49.71	6.731	51.47
		(2.259)		(1.531)		(1.556)	
	σ_e^2	3.083	22.41	3.099	24.27	3.139	24.01
		(0.523)		(0.585)		(0.600)	
	h^2	0.132		0.260		0.245	
WW	σ_a^2	139.230	19.45	231.475	22.35	209.009	19.83
		(52.205)		(84.339)		(75.127)	
	σ_{cg}^2	297.158	41.50	465.301	44.94	488.131	46.31
		(95.332)		(139.124)		(140.685)	
	σ_e^2	279.626	39.05	338.71	32.71	356.821	33.86
		(49.849)		(68.278)		(65.268)	
	h^2	0.194		0.223		0.198	
YW	σ_a^2	430.111	22.28	612.485	23.16	538.373	19.89
		(176.121)		(220.039)		(198.358)	

σ_{cg}^2	628.619 (212.535)	32.56	1097.472 (365.987)	41.50	1183.183 (378.962)	43.72
σ_e^2	871.793 (169.503)	45.16	934.302 (194.968)	35.33	984.992 (185.882)	36.39
h^2	0.222		0.231		0.198	

369

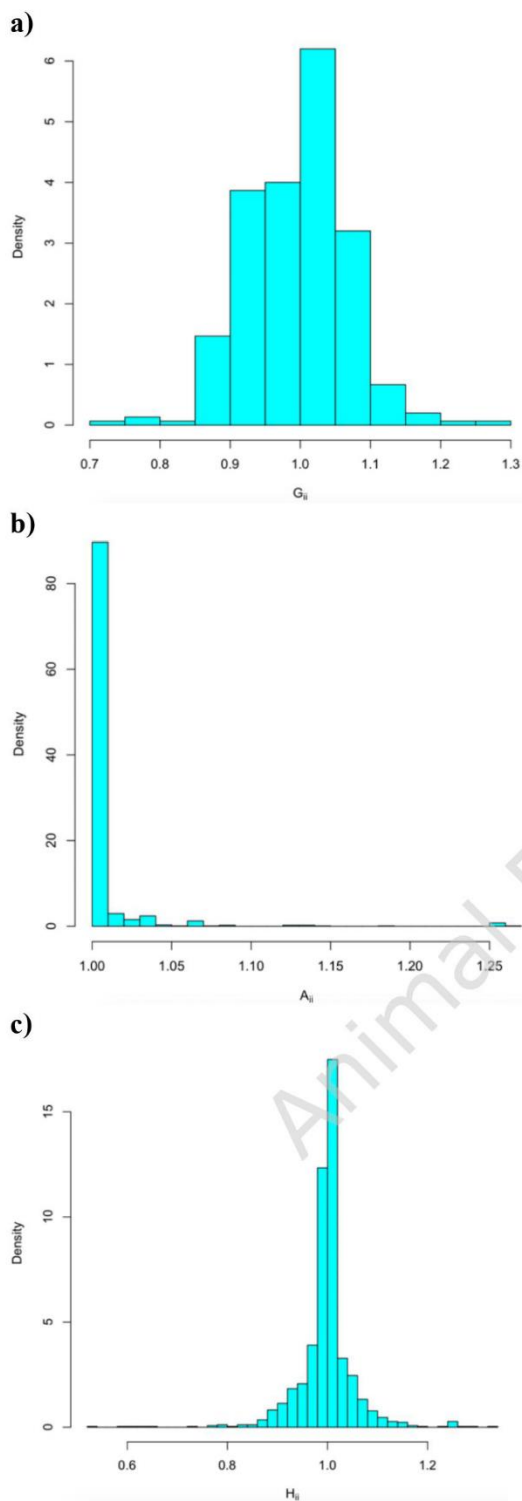
370

Animal Bioscience

Table 2

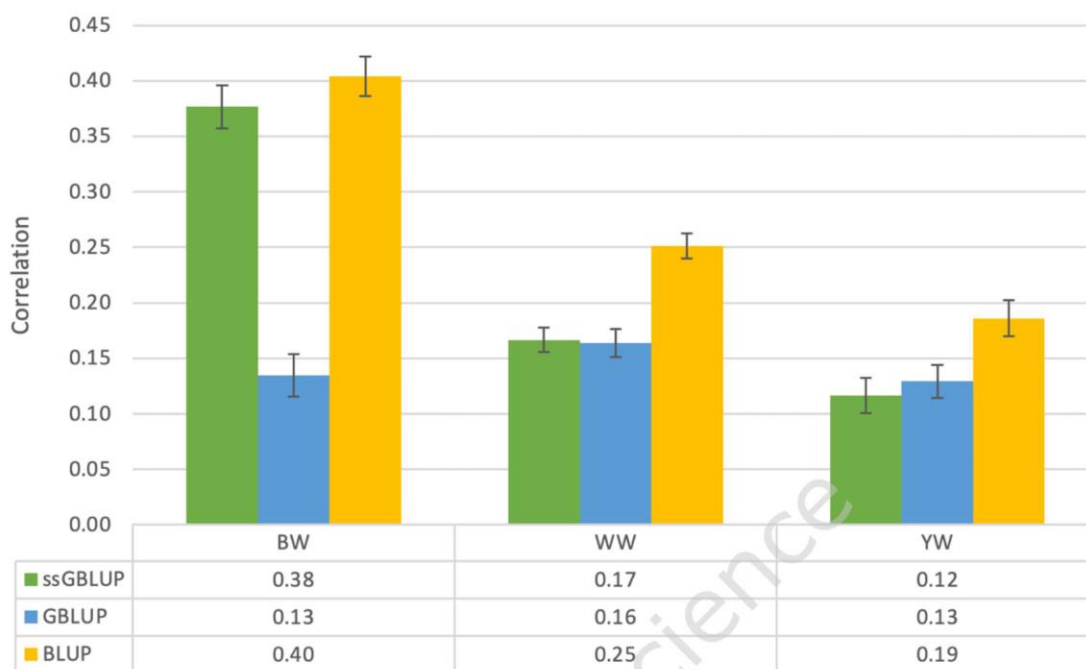
Pearson correlation estimates between corrected phenotypes and estimated breeding values obtained from a validation with 100 random partitions when fitting (BLUP), genomic BLUP (GBLUP) and single-step GBLUP (ssGBLUP) models, for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a Mexican Braunvieh cattle population. Standard deviations are in parenthesis.

	BW	WW	YW
GBLUP	0.135 (0.192)	0.164 (0.126)	0.129 (0.149)
BLUP	0.404 (0.180)	0.251 (0.114)	0.186 (0.153)
ssGBLUP	0.377 (0.193)	0.167 (0.111)	0.117 (0.160)



379 **Figure 1.** Distribution of the diagonal entries of a) Genomic relationship matrix derived from markers,
 380 **G**, b) the additive genomic relationship matrix derived from pedigree, **A**, and c) the combined matrix,
 381 **H**.

382



383

384 **Figure 2.** Pearson correlation coefficients between the observed and corrected phenotypes obtained
 385 with best linear unbiased predictor (BLUP), genomic BLUP (GBLUP), and single-step GBLUP
 386 (ssGBLUP) methods for birth weight (BW), weaning weight (WW), and yearling weight (YW) of a
 387 Mexican Braunvieh cattle population.

4. REGRESIÓN CUANTIL PARA PREDICCIÓN DE CARACTERES COMPLEJOS EN BOVINOS SUIZO EUROPEO USANDO MARCADORES SNP Y PEDIGRÍ

Jonathan Emanuel Valerio-Hernández¹, Agustín Ruíz-Flores¹, Paulino Pérez-
Rodríguez²

¹ Departamento de Zootecnia, Universidad Autónoma Chapingo, Km 38.5
Carretera México-Texcoco. Chapingo, Estado de México. CP 56227

² Programa de Estadística, Colegio de Postgraduados. Montecillo, Estado de
México. CP 56264

Artículo enviado a la *Revista Mexicana de Ciencias Pecuarias*

1 **REGRESIÓN CUANTIL PARA PREDICCIÓN DE CARACTERES COMPLEJOS EN**
2 **BOVINOS SUIZO EUROPEO USANDO MARCADORES SNP Y PEDIGRÍ**
3 **QUANTILE REGRESSION FOR PREDICTION OF COMPLEX TRAITS IN**
4 **BRAUNVIEH CATTLE USING SNP MARKERS AND PEDIGREE**

5 Jonathan Emanuel Valerio-Hernández¹, Paulino Pérez-Rodríguez^{2*}, Agustín Ruíz-Flores¹

6 ¹ *Posgrado en Producción Animal, Universidad Autónoma Chapingo, Carretera Federal*

7 *México-Texcoco Km 38.5, 56227, Texcoco, Estado de México, México.*

8 ² *Socio Economía Estadística e Informática, Colegio de Postgraduados, Carretera Federal*

9 *México-Texcoco Km 36.5, 56230, Texcoco, Estado de México.*

10

11 *Autor para correspondencia, e-mail: perpdgo@colpos.mx

12 **Resumen**

13 La predicción genómica generalmente supone variables aleatorias independientes e
14 idénticamente distribuidas con media cero e igual varianza. Esto no siempre se cumple,
15 además puede haber fenotipos distantes de la media poblacional, los que usualmente se
16 eliminan al realizar la predicción. La regresión cuantil (QR) afronta aspectos estadísticos
17 como alta dimensionalidad, multicolinealidad y distribución fenotípica diferente de la
18 normal. El objetivo de este trabajo fue comparar QR utilizando información de marcadores
19 y pedigrí con los métodos alternativos tales como mejor predicción lineal insesgada
20 genómica (GBLUP) y mejor predicción lineal insesgada genómica en una etapa (ssGBLUP)
21 para analizar los pesos al nacimiento (PN), destete (PD) y al año (PA) de bovinos Suizo
22 Europeo y datos simulados con diferente grado de asimetría y proporción de datos atípicos.
23 La capacidad predictiva de los modelos se evaluó mediante validación cruzada. El
24 desempeño predictivo de QR tanto solo con información de marcadores como con
25 marcadores más pedigrí, con el conjunto de datos reales, fue mejor que las metodologías
26 GBLUP y ssGBLUP para PD y PA. Para PN GBLUP y ssGBLUP fueron mejores, sin
27 embargo, solo se utilizaron los cuantiles 0.25, 0.50 y 0.75, y la distribución de PN no fue
28 asimétrica. En el experimento de datos simulados, las correlaciones entre efectos de marcador
29 “verdadero” y efectos estimados, así como las correlaciones de señales “verdaderas” y
30 estimadas fueron más altas cuando se usó QR comparado con GBLUP. Las ventajas de QR
31 fueron más notorias con distribución asimétrica de los fenotipos y con mayor proporción de
32 datos atípicos, como fue el caso del conjunto de datos simulados.

33

34 **Palabras clave:** Regresión cuantil, GBLUP, ssGBLUP

QUANTILE REGRESSION FOR PREDICTION OF COMPLEX TRAITS IN BRAUNVIEH CATTLE USING SNP MARKERS AND PEDIGREE

Summary

Genomic prediction models usually assume random and independent variables with zero mean and equal variance. This is seldomly met, additionally, there could be phenotypes far away from the mean, they are usually discarded in these evaluations. The quantile regression (QR) tackles statistic issues such as high dimensionality, multicollinearity and phenotypes that do not follow a normal distribution. The objective of this work was to compare QR using marker and pedigree information with the alternative methods such as genomic best linear unbiased prediction (GBLUP) and single-step genomic best linear unbiased prediction (ssGBLUP). The traits analyzed were birth (BW), weaning (WW), and yearling (YW) weights of Braunvieh cattle as well as simulated data with different degree of asymmetry and proportion of atypical data. The predictive ability of the models was evaluated through cross-validation. The predictive performance of QR both with only marker and with marker and pedigree information, with the real dataset was better than GBLUP and ssGBLUP for WW and YW. For BW, GBLUP and ssGBLUP were better, however, only the 0.25, 0.50 and 0.75 quantiles were used, and the distribution of BW was not asymmetric. In the experiment with simulated data, the correlation estimates between the “true” marker effects and the estimated effects, as well as the correlation estimates between “true” and estimated signals were higher when QR was used compared with GBLUP. The advantages of QR were more notorious with asymmetric distribution of the phenotypes and with greater proportion of atypic data, as was the case for the simulated dataset.

Keywords: Quantile regression, GBLUP, ssGBLUP

58 **Introducción**

59 La motivación principal del método de regresión cuantil (QR) reside en que la mayoría de
60 los modelos para evaluación genética asumen normalidad, lo que no siempre se cumple. Con
61 el método QR se obtienen resultados más robustos y una visión más amplia de las variables
62 explicativas sobre las dependientes⁽¹⁾. En mejoramiento genético siempre ha habido interés
63 en el extremo de las distribuciones de características importantes para detectar posibles
64 señales genéticas⁽²⁾. Los datos generados a partir de experimentos ómicos frecuentemente
65 son complejos y de gran dimensión, por consiguiente existe un desafío estadístico para
66 extraer información biológica relevante del gran volumen de datos^(2,3). El modelo de
67 regresión normal mide el impacto de covariables en la media de la distribución, en cambio
68 QR puede dar una visión amplia de los datos al probar la relación de diferentes cuantiles de
69 la variable de respuesta y covariables, de tal forma que se pueden estudiar las colas de
70 distribuciones (e.g. pesos bajos y altos).

71 Otra característica de QR para los genetistas es la solidez de los supuestos de distribución y
72 las observaciones atípicas; cualquier desarrollo de nueva tecnología de alto rendimiento
73 siempre viene con errores de medición y sesgos sistemáticos a menudo difíciles de controlar.
74 El uso de un enfoque sólido como QR hace que la inferencia sea menos sesgada y esté menos
75 sujeta a falsos positivos⁽²⁾. En estudios recientes que utilizan QR, se describen aplicaciones
76 diversas como estudios de asociación genética⁽⁴⁾, genética de poblaciones⁽⁵⁾, expresión
77 génica^(6,7) y selección genómica^(8,9,10).

78 Uno de los primeros estudios donde se utilizó QR para predecir el mérito genético individual
79 lo presentaron Nascimento et al.⁽¹¹⁾, quienes utilizaron datos simulados encontrando ventajas
80 al usar QR frente a metodologías convencionales. El mismo año, Barroso et al.⁽¹²⁾ publicaron
81 resultados usando QR para ajustar curvas de crecimiento con datos de cerdos y marcadores

moleculares; no solo ajustaron con éxito las curvas de crecimiento sino que identificaron marcadores de importancia asociados con la característica estudiada. Posteriormente, Mendes dos Santos et al.⁽¹³⁾ predijeron el mérito genético en cerdos de características con densidades asimétricas usando QR, encontraron que la exactitud de las predicciones de QR fueron mejores que las del método LASSO Bayesiano. Otro trabajo similar del mismo equipo de investigadores fue presentado por Nascimento et al.⁽¹⁴⁾, pero con datos de frijol. Recientemente Pérez-Rodríguez et al.⁽¹⁰⁾ extendieron el modelo de regresión cuantil para incluir información de pedigrí a través del uso de la matriz de relaciones genéticas aditivas, mejorando aún más la habilidad predictiva de los modelos y al mismo tiempo identificando las proporciones de las varianzas atribuidas a marcadores, relaciones entre individuos y el residual, lo cual permite obtener de manera precisa una partición de la varianza fenotípica. El modelo desarrollado fue aplicado en genética vegetal con datos de trigo.

En este trabajo se utilizaron pesos al nacimiento, destete y al año de bovinos Suizo Europeo para comparar los modelos: 1) QR con información de marcadores moleculares SNP (QRM), 2) QR incluyendo simultáneamente información de marcadores moleculares y genealógica derivada del pedigrí (QRH); 3) GBLUP el cual al igual que QRM solo incluyó información de marcadores moleculares, y 4) evaluación genómica en un solo paso (ssGBLUP) que incluyó información de marcadores y pedigrí. El desempeño predictivo de los modelos propuestos se evaluó mediante validación cruzada y utilizando información simulada.

Materiales y métodos

Información utilizada

Genotipos

Se utilizó información de 300 animales (236 hembras, 64 machos) nacidos de 2001 a 2016 en cinco hatos ubicados en el Este, Centro y Oeste de México. Se recolectaron muestras de

pelo para su genotipado por la compañía GeneSeek (Lincoln, <https://www.neogen.com/>, NE, EE. UU.), utilizando el panel GeneSeek® Genomic Proler Bovine LDv.4, con 30000 y 50000 marcadores SNP, 150 animales con cada Chip. Se utilizaron los SNP en común entre los chips de 30K y 50K (12,835 SNP). Los genotipos se recodificaron como AA = 0, AB = 1 y BB = 2. Los genotipos de marcadores faltantes se imputaron al azar, generando muestras de la distribución $Binomial(2, \hat{p})$ donde $\hat{p}$ es la frecuencia del alelo mayor, calculado a partir de los genotipos de marcadores observados. Se eliminaron los marcadores monomórficos o aquellos con frecuencia de alelos menores (MAF) menor que 0.04. Después del control de calidad, 9,628 de los 12,835 SNP en común entre los dos chips estuvieron disponibles para análisis adicionales.

Fenotipos

La información fenotípica y de pedigrí de la población de ganado Suizo Europeo se obtuvo de la base de datos de la Asociación Mexicana de Criadores de Ganado Suizo de Registro. Los registros de pesos al nacer (PN), al destete (PD) y al año (PA) se utilizaron para el análisis. La edición de los fenotipos fue similar para PN, PD y PA, se descartaron los registros de animales no relacionados genéticamente con aquellos genotipados o sin información referente a hato, edad de la madre y manejo. Los grupos contemporáneos (GC) se definieron al eliminar los animales en GC de 2 individuos o con varianza igual a cero. Para PN los GC se definieron combinando los efectos del hato (8 hatos), año (1998 a 2016) y temporada de nacimiento; las temporadas de nacimiento se definieron considerando el calendario juliano, de 80 a 171 d, primavera; de 172 a 264 d, verano; de 265 a 354 d, otoño; de 355 a 366 d y de 1 a 79 d, invierno. Después de la edición de datos para PN se obtuvieron 330 registros. Para PD y PA los GC se definieron combinando los efectos del hato (6 hatos), año (de 1998 a 2016), temporada de nacimiento (igual que PN) y manejo. En el caso de PD los grupos de

manejo se definieron de tres formas: terneros alimentados con leche de su madre; leche de su madre más alimento balanceado; y leche de su madre y nodriza más alimentación balanceada. Para PA los grupos de manejo se definieron de tres maneras: animales en pastoreo; animales en pastoreo con concentrado alimenticio; y animales estabulados con alimentación equilibrada. La edición de datos de PD y PA finalizó con 267 y 232 registros para análisis posteriores. En el Cuadro 1 se presenta un resumen del número de animales genotipados, genotipados y fenotipados para PN, PD y PA. La Figura 1 muestra las gráficas de violín para PN, PD y PA.

Modelos

Modelo de regresión cuantil con marcadores (QRM)

El modelo para regresión cuantil es:

$$y_i = \mu + \mathbf{x}_i^t \boldsymbol{\beta} + w_i,$$

donde y_i es el valor del fenotipo del i -ésimo animal; μ es un intercepto; $\mathbf{x}_i^t = (x_{i1}, \dots, x_{ip})$ representa la i -ésima fila de la matriz de marcadores, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ es el vector de coeficientes de regresión asociados con marcadores y w_i son variables aleatorias independientes tales que su cuantil $\theta \in (0,1)$ es cero. La estimación de los coeficientes de regresión para un cuantil de interés θ fijo se obtiene al resolver el problema de minimización siguiente:

$$\min\{\sum_{i=1}^n \rho_{\theta}(y_i - \mu - \mathbf{x}_i^t \boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j|\},$$

donde $\sum_{j=1}^p |\beta_j|$ es la suma de los valores absolutos de los coeficientes de regresión; λ es el parámetro de penalización que controla la intensidad de la regularización; y $\rho_{\theta}(\cdot)$ es la función definida como⁽¹⁾:

$$\rho_{\theta}(t_i) = \begin{cases} \tau \times t_i & \text{Si } t_i \geq 0 \\ -(1 - \tau) \times t_i & \text{Si } t_i < 0, \end{cases}$$

donde $t_i = y_i - \mu - \mathbf{x}_i^t \boldsymbol{\beta}$. Después de estimar los parámetros del modelo, los valores de cría estimados mediante marcadores (GEVB) se obtienen mediante la siguiente expresión:

$$GEVB(\tau) = \hat{y}_i(\tau) = \sum_{j=1}^p x_{ij} \hat{\beta}_j(\tau),$$

donde $\hat{\beta}_j(\tau)$ es el efecto del j -ésimo marcador, definido por la relación funcional obtenida para el cuantil de interés.

El modelo QR puede extenderse para incluir otros términos, en particular para las características de crecimiento se utiliza el siguiente modelo:

$$y_i = \mu + \mathbf{s}_i^t \boldsymbol{\zeta} + \mathbf{c}_i^t \boldsymbol{\varrho} + \mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{w}_i,$$

donde y_i es el valor del fenotipo de la característica analizada (PN, PD o PA) del i -ésimo animal, μ es un intercepto; $\mathbf{s}_i^t = (s_{i1}, \dots, s_{if})$ la i -ésima fila de la matriz de incidencia para los efectos fijos (sexo, edad de la madre, manejo), $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_f)^t$ los coeficientes de regresión para los efectos fijos, $\mathbf{c}_i^t = (c_{i1}, \dots, c_{it})$ la i -ésima fila de la matriz de incidencia para efectos aleatorios de grupo contemporáneo (54, 43 y 37 para PN, PD and PA), $\boldsymbol{\varrho} = (\varrho_1, \dots, \varrho_t)^t$ efectos aleatorios de grupo contemporáneo, el resto de los términos como se describieron anteriormente.

GBLUP

El modelo está dado por:

$$y_i = \mu + \mathbf{s}_i^t \boldsymbol{\zeta} + \mathbf{c}_i^t \boldsymbol{\varrho} + \mathbf{z}_i^t \mathbf{u} + e_i,$$

donde $\mathbf{z}_i^t = (z_{i1}, \dots, z_{in})$ es la i -ésima fila de la matriz que conecta fenotipos con genotipos, $\mathbf{u} = (u_1, \dots, u_n)^t$ es el vector de efectos aleatorios para animales. Las varianzas genéticas aditiva, de grupo contemporáneo y residual se asumen $Var(\mathbf{u}) = \mathbf{G}\sigma_u^2$, $Var(\mathbf{c}) = \mathbf{I}\sigma_{cg}^2$, y

172 $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$, respectivamente. La matriz de relaciones genómicas, $\mathbf{G}$, se calcula como lo
 173 describen López-Cruz et al.⁽¹⁵⁾ y Pérez-Rodríguez et al.⁽¹⁶⁾; brevemente, $\mathbf{G} = \mathbf{W}\mathbf{W}'/p$, donde
 174 $\mathbf{W}$ es la matriz de marcadores estandarizada y centrada (cada marcador centrado restando la
 175 media de frecuencia de alelo y estandarizado, dividiendo por la desviación estándar de la
 176 muestra de la frecuencia de alelos), p es el número total de marcadores, e_i variables aleatorias
 177 normales e independientes con distribución normal con media 0 y varianza σ_e^2 .

178 **Modelo de regresión cuantil en un solo paso (QRH)**

179 Este método se considera una extensión del modelo de cuantiles para una matriz de relaciones
 180 construida utilizando matrices de relaciones para animales genotipados y no genotipados y
 181 de los cuales se tiene disponible un pedigrí. La matriz que resulta se conoce en la literatura
 182 como matriz $\mathbf{H}^{(17,18)}$, esta matriz está dada por:

$$183 \quad \mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_a^{-1} - \mathbf{A}_{gg}^{-1} \end{bmatrix},$$

184 donde, $\mathbf{A}_{gg}$ es una submatriz de $\mathbf{A}$ para animales genotipados, $\mathbf{G}_a = \beta\mathbf{G} + \alpha$; β y α se obtienen
 185 al resolver el sistema de ecuaciones:

$$186 \quad \begin{cases} Avg(diag(\mathbf{G}))\beta + \alpha = Avg(diag(\mathbf{A}_{gg})) \\ Avg(\mathbf{G})\beta + \alpha = Avg(\mathbf{A}_{gg}) \end{cases}.$$

187 El modelo QRH está dado por:

$$188 \quad y_i = \mu + \mathbf{s}_i^t \boldsymbol{\zeta} + \mathbf{c}_i^t \boldsymbol{\varrho} + \mathbf{z}_i^t \mathbf{u} + w_i,$$

189 donde $Var(\mathbf{u}) = \sigma_H^2 \mathbf{H}$, el resto de los términos como se describieron previamente.

190 **Modelo de regresión GBLUP en un solo paso (ssGBLUP)**

191 El modelo ssGBLUP es equivalente al modelo GBLUP descrito anteriormente con la
 192 diferencia de que se reemplaza la matriz de relaciones genómicas $\mathbf{G}$ por la matriz de
 193 relaciones genéticas extendida $\mathbf{H}$, se asume que $Var(\mathbf{u}) = \mathbf{H}\sigma_H^2$.

194 Validación cruzada

195 La capacidad predictiva de los modelos se evaluó mediante validación cruzada, la cual se
 196 realizó de la siguiente manera. El conjunto de datos se dividió en 5 grupos del mismo tamaño
 197 $\{S_1, S_2, \dots, S_5\}$, 80% de los datos fueron usados para entrenamiento del modelo, el 20%
 198 restante para validación. Por ejemplo, $\{S_2\}$ se usa como grupo de validación y el conjunto
 199 $\{S_1, S_3, \dots, S_5\}$ para entrenamiento del modelo. Los modelos se ajustaron utilizando el
 200 conjunto de entrenamiento, y el modelo ajustado se utilizó para obtener predicciones para el
 201 conjunto de validación. Este procedimiento se repitió cinco veces y se obtuvieron
 202 predicciones para cada grupo. Las correlaciones entre fenotipos observados y predichos se
 203 calcularon y promediaron para los conjuntos de prueba⁽¹⁹⁾.

204 Simulación

205 Con la finalidad de evaluar el poder predictivo del modelo QR frente a GBLUP se realizó
 206 además una simulación de datos asimétricos con presencia de datos atípicos; la simulación
 207 del presente trabajo es análoga a la utilizada por Pérez-Rodríguez et al.⁽¹⁰⁾. Se consideraron
 208 los 9,628 SNPs resultantes del control de calidad descrito anteriormente para 300 animales,
 209 la simulación de los datos se realizó considerando el modelo lineal:

$$210 \quad y_i = \mu + \sum_{j=1}^{9,628} x_{ij}\beta_j + e_i,$$

211 donde $i = 1, \dots, 300$, con $\mu = 39$ para PN, se supuso que los errores vienen de una
 212 distribución normal sesgada (SN_c) con media 0, varianza 1 e índice de asimetría γ_1 , es decir
 213 $e_i \sim SN_c(0, \sigma, \gamma_1)$, con $\sigma = \sqrt{1 - h^2}$, se consideró h^2 con un valor de 0.35, $\gamma_1 =$

$$214 \quad \sqrt{\frac{2}{\pi}} \rho^3 \left(\frac{4}{\pi} - 1 \right) \left(1 - \frac{2\rho^2}{\pi} \right)^{-3/2}, \rho \in \{0.950, 0.975, 0.999\} \text{ lo que lleva a diferentes grados de}$$

215 sesgo positivo. Se fijaron 50 marcadores con efecto diferente de cero simulándolos de una de
 216 distribución normal con media 0 y varianza $\sqrt{1 - h^2}/50$, el resto de los marcadores se

establecieron en 0; las posiciones de los marcadores muestreados se tomaron al azar. Para introducir datos atípicos en los fenotipos, se generó cierta proporción de los residuos de $e_i \sim SN_C(0, 3, \gamma_1)$ al azar, se consideraron dos proporciones, 5 y 10%, por lo que se tomaron muestras de una mezcla de dos componentes de distribuciones normales sesgadas. Se generaron 6 conjuntos de datos, tres diferentes coeficientes de asimetría 0.950, 0.975, 0.999 con sus dos alternativas de proporción de datos atípicos 5% y 10%; y se ajustó el modelo QRM con $\theta = \{0.25, 0.50, 0.75\}$ para compararlo con GBLUP. Para cada análisis se calculó la correlación entre los β 's verdaderos y estimados, la correlación entre las señales verdaderas $X\beta$ y estimadas $X\hat{\beta}$ y el componente de varianza asociado con los residuos para cada modelo, el cual es una forma de evaluar la bondad de ajuste de los modelos. También se consideró el Criterio de Información de Desviación (DIC), éste se puede usar para seleccionar modelos candidatos; los modelos con menor DIC se prefieren contra modelos de mayor DIC^(10,20).

Software y ajuste de los modelos

Los modelos de regresión por cuantiles se ajustaron usando una estrategia computacional similar a la descrita por Pérez-Rodríguez et al.⁽¹⁰⁾. Las adaptaciones de los algoritmos para incluir efectos fijos y aleatorias no presentan gran dificultad computacional. Los códigos para el ajuste de los modelos se desarrollaron en los lenguajes de programación R⁽²¹⁾ y C. En todos se seleccionaron tres cuantiles, $\theta = \{0.25, 0.50, 0.75\}$. Los modelos GBLUP y ssGBLUP se ajustaron con la librería de R BGLR⁽²²⁾.

237 **Resultados**

238 *Información real*

239 El Cuadro 2 muestra los resultados del experimento realizado con datos de PN, PD y PA de
 240 una población de bovinos Suizo Europeo, evaluados bajo dos escenarios 1) solo con
 241 información de marcadores, y 2) información de marcadores y pedigrí. De manera general
 242 las correlaciones entre valores observados y predichos más altas se obtuvieron con QR,
 243 excepto para PN donde las correlaciones de GBLUP y ssGBLUP fueron más altas que las
 244 obtenidas con QRM y QRH, sin embargo, las correlaciones de QRM $\theta = 0.75$ y QRH $\theta =$
 245 0.75 fueron cercanas a las obtenidas con GBLUP y ssGBLUP (0.7902 vs 0.7924), (0.6981
 246 vs 0.7055), respectivamente. Los valores de MSE más bajos se obtuvieron con QRM $\theta =$
 247 0.75 y QRH $\theta = 0.75$ en la característica de PD, mientras que en las características PN y PA
 248 los valores más bajos se obtuvieron con GBLUP y ssGBLUP. Los componentes de varianza
 249 asociados con el error obtenidos con QRM y QRH fueron menores que los obtenidos con
 250 GBLUP y ssGBLUP. Los valores de DIC más bajos de manera general se obtuvieron con
 251 QRM $\theta = 0.75$ y QRH $\theta = 0.75$, excepto para PN con el escenario de solo marcadores,
 252 donde el DIC más bajo se obtuvo con QRM $\theta = 0.25$.

253 *Información simulada*

254 Los resultados del ejercicio de simulación donde se compara QR con GBLUP bajo diferentes
 255 grados de asimetría y proporciones de datos atípicos se muestran en el Cuadro 3. En la
 256 columna 2 se registran las correlaciones entre los efectos de marcador “verdaderos” y los
 257 efectos de marcador estimados, las correlaciones obtenidas con QR fueron más altas que las
 258 obtenidas con GBLUP. La columna 3 muestra las correlaciones entre las “señales
 259 verdaderas” y las estimadas, las correlaciones más altas se obtuvieron con QR. La columna

4 registra la estimación de los componentes de varianza asociados al error y la columna 5 los valores de DIC, los valores más bajos en ambas columnas se obtuvieron con QR $\theta = 0.75$.

Discusión y conclusiones

En este estudio se compararon las metodologías de análisis QR con GBLUP y ssGBLUP. Esta comparación se hizo con fenotipos simulados con diferentes grados de asimetría y proporciones de datos atípicos y datos reales de pesos al nacimiento, al destete y al año.

Información real

Las correlaciones de fenotipos observados y predichos obtenidos de la validación cruzada con datos reales fueron más altas al usar QRM y QRH en las características de PD y PA. Para PN las correlaciones más altas se obtuvieron con GBLUP y ssGBLUP; sin embargo, en este estudio solo se probaron tres cuantiles 0.25, 0.50 y 0.75; es probable que al probar otros cuantiles las correlaciones sean mayores que GBLUP y ssGBLUP, hay evidencia en otros estudios donde QR es mejor que GBLUP como en el caso del trabajo de Nascimento et al.⁽¹⁴⁾, quienes compararon QR con modelos como con BLASSO, BayesB y RR-BLUP. Estos autores encontraron una ganancia en la capacidad predictiva de QR frente a RR-BLUP del 15.15%, cabe señalar que matemáticamente RR-es equivalente a GBLUP, además de que los conjuntos de datos usados en este experimento presentaban asimetría.

Los valores del cuadrado medio del error (MSE) miden el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima, por lo que se prefieren valores bajos; los promedios de MSE de QRM y QRH fueron menores que los obtenidos con GBLUP y ssGBLUP únicamente para PD. El estimador de la varianza residual es un indicativo de que tan bien o mal se ajusta el modelo a los datos observados, se prefieren valores bajos; los componentes de varianza del error más bajos se obtuvieron con QRM y QRH para las tres características analizadas. Finalmente, el valor de DIC se usa para

seleccionar modelos candidatos y al igual que MSE y los componentes de varianza del error se prefieren valores bajos. Los valores de DIC más bajos se obtuvieron con QRM $\theta = 0.75$ y QRH $\theta = 0.75$, excepto en el escenario de solo marcadores y PN, donde el DIC más bajo se obtuvo con QRM $\theta = 0.25$. El cuadrado medio del error, la varianza residual y el DIC son valores que ayuda a escoger el modelo de mejor ajuste. Al examinar estos valores de manera conjunta se observa que QRM y QRH son mejores en algunos de ellos mientras que en otros no, es decir, QR presenta un desempeño igual o superior a GBLUP y ssGBLUP; aunque debe resaltarse que solo se probaron tres cuantiles y que QR presenta ventajas al usarse en datos asimétricos y datos atípicos, para este caso solo había datos atípicos y las distribuciones no presentaban asimetría. Mendes dos Santos et al.⁽¹³⁾ compararon QR con el método bayesiano del lazo (BLASSO), estos autores reportaron un aumento de 6.7 y 20.0% en la precisión y consideraron los cuantiles 0.15 y 0.45 en la evaluación del rendimiento de la canal y el grosor del tocino, respectivamente, sin embargo, las características evaluadas en su estudio fueron asimétricas.

Información simulada

En el experimento de datos simulados las correlaciones entre efectos de marcador “verdadero” y efectos estimados, así como las correlaciones de señales “verdaderas” y estimadas fueron más altas cuando se usó QR comparado con GBLUP. Estos resultados son similares a los obtenidos por Pérez-Rodríguez et al.⁽¹⁰⁾, quienes simularon datos con tres diferentes coeficientes de asimetría 0.75, 0.95, 0.999 con 5% y 10% de datos atípicos y encontraron que las correlaciones obtenidas con QR fueron más altas que las obtenidas con la regresión de cresta bayesiana (BRR), además, este patrón fue más evidente con una mayor asimetría y proporción de datos atípicos. En este trabajo se realizaron simulaciones con coeficientes de asimetría de 0.950, 0.975, 0.999 y los cuantiles con los que se obtuvo

308 correlaciones más altas variaron entre 0.25 y 0.50; la ventaja de QR es que se puede probar
309 diferentes cuantiles obteniendo mejores resultados dependiendo del cuantil utilizado, esta
310 ventaja en la capacidad de predicción de efectos de marcadores y señales ha sido aprovechada
311 por otros investigadores como Nascimento et al.⁽⁴⁾, quienes no encontraron ninguna
312 asociación de rasgo usando el modelo tradicional GWAS de SNP único, pero, cuando usaron
313 QR con cuantiles extremos como 0.1 el modelo fue capaz de encontrar hasta 7 SNPs
314 asociados a las características estudiadas.

315 Los coeficientes de varianza del error indican qué tan bien se ajusta el modelo propuesto a
316 los datos estudiados, entre más pequeño sea este valor, el ajuste es mejor, el DIC también es
317 otro valor que se usa para comparar modelos candidatos. Los modelos con un DIC más
318 pequeño se prefieren a los modelos con un DIC mayor⁽²⁰⁾. Los estimadores de varianza
319 residual y los valores de DIC más bajos se obtuvieron con QR $\theta = 0.75$. quizá esto se deba
320 a que en la simulación se usaron coeficientes de asimetría altos 0.950, 0.975, 0.999, entonces
321 por consiguiente se espera que un cuantil que se ajuste mejor sea el más alto, en este caso
322 0.75.

323 QR tuvo un desempeño igual o mejor que GBLUP y ssGBLUP para predecir características
324 de crecimiento PN, PD y PA, las ventajas de este método son más notorias cuando los datos
325 están más sesgados y presentan mayor proporción de datos atípicos como en el caso del
326 experimento de simulación.

327 Agradecimientos: Los autores agradecen al Consejo Nacional de Ciencia y Tecnología,
328 México, por el apoyo financiero para el primer autor durante sus estudios de doctorado. Los
329 autores también agradecen a la Asociación Mexicana de Criadores de Ganado Suizo de
330 Registro por permitir el uso de sus bases de datos, a los criadores cooperantes por su amable
331 cooperación en este estudio.

332 Conflictos de interés: Los autores declaran que no existen conflictos de interés.

333

Literatura citada

1. Koenker R, Bassett G. Regression quantiles. *Econometrica* 1978;46(1):33–50.
<https://doi.org/10.2307/1913643>
2. Briollais L, Durrieu G. Application of quantile regression to recent genetic and -omic studies. *Hum Genet* 2014;133(8):951-966. <https://doi.org/10.1007/s00439-014-1440-6>
3. Wang L, Wu Y, Li R. Quantile regression for analyzing heterogeneity in ultra-high dimension. *J Am Stat Assoc* 2012;107(497):214-222.
<https://doi.org/10.1080/01621459.2012.656014>
4. Nascimento M, Nascimento ACC, Silva FF, Barili LD, Do Vale NM, Carneiro JE, Cruz CD, Carneiro PCS, Serão NVL. Quantile regression for genome-wide association study of flowering time-related traits in common bean. *PLoS One* 2018;13(1):1-14.
<https://doi.org/10.1371/journal.pone.0190303>
5. Fisher E, Schweiger R, Rosset S. Efficient construction of test inversion confidence intervals using quantile regression. *J Comput Graph Stat* 2016;29:140-148,
<http://arxiv.org/abs/1612.02300>
6. Logan JAR, Petrill SA, Hart SA, Schatschneider C, Thompson LA, Deater-Deckard K, De Thorne LS, Bartlett C. Heritability across the distribution: An application of quantile regression. *Behav Genet* 2012;42(2):256–267. <https://doi.org/10.1007/s10519-011-9497-7>
7. Vinciotti V, Yu K. M-quantile regression analysis of temporal gene expression data. *Stat Appl Genet Mol Biol* 2009;8(1). <https://doi.org/10.2202/1544-6115.1452>
8. Gianola D, Cecchinato A, Naya H, Schön CC. Prediction of complex traits: Robust alternatives to best linear unbiased prediction. *Front Genet* 2018;9:195.
<https://doi.org/10.3389/fgene.2018.00195>

- 358 9. Oliveira GF, Nascimento ACC, Nascimento M, Sant'Anna IdeC, Romero JV, Azevedo
 359 CF, Bhering LL, Moura ETC. Quantile regression in genomic selection for oligogenic
 360 traits in autogamous plants: A simulation study. PLoS One 2021;16(1):1-12.
 361 <https://doi.org/10.1371/journal.pone.0243666>
- 362 10. Pérez-Rodríguez P, Montesinos-López OA, Montesinos-López A, Crossa J. Bayesian
 363 regularized quantile regression: A robust alternative for genome-based prediction of
 364 skewed data. Crop J 2020;8(5):713-722. <https://doi.org/10.1016/j.cj.2020.04.009>
- 365 11. Nascimento M, e Silva FF, de Resende MDV, Cruz CD, Nascimento ACC, Viana JMS,
 366 Azebedo CF, Barroso LMA. Regularized quantile regression applied to genome-
 367 enabled prediction of quantitative traits. Genet Mol Res 2017;16(1).
 368 <https://doi.org/10.4238/GMR16019538>
- 369 12. Barroso LMA, Nascimento M, Nascimento ACC, Silva FF, Serão NVL, Cruz, CD,
 370 Resende MDV, Silva FL, Azevedo CF, Lopes PS, Guimarães SEF. Regularized quantile
 371 regression for SNP marker estimation of pig growth curves, J. Animal Sci Biotechnol
 372 2017;8:59. <https://doi.org/10.1186/s40104-017-0187-z>
- 373 13. Mendes dos Santos P, Nascimento ACC, Nascimento M, Fonseca e Silva F, Azevedo CF,
 374 Mota RR, Guimarães SEF, Lopes PS. Use of regularized quantile regression to predict
 375 the genetic merit of pigs for asymmetric carcass traits. Pesqui Agropecu Bras
 376 2018;53(9):1011–1017. <https://doi.org/10.1590/S0100-204X2018000900004>
- 377 14. Nascimento AC, Nascimento M, Azevedo C, Silva F, Barili L, Vale N, Carneiro J, Cruz
 378 C, Carneiro PC, Serão N. Quantile regression applied to genome-enabled prediction of
 379 traits related to flowering time in the common bean. Agronomy 2019;9(12):1-11.
 380 <https://doi.org/10.3390/agronomy9120796>
- 381 15. López-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink J-L, Singh RP,

- 382 Autrique E, de los Campos G. Increased prediction accuracy in wheat breeding trials
 383 using a Marker $\times$ Environment interaction genomic selection model. *G3 Genes Genoms*
 384 *Genetics* 2015;5(4):569–582. <https://doi.org/10.1534/g3.114.016097>
- 385 16. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, Autrique E,
 386 Campos G, Burgueño J, Dreisigacker S. Single-step genomic and Pedigree Genotype $\times$
 387 Environment interaction models for predicting wheat lines in international
 388 environments. *Plant Genome* 2017;10(2).
 389 <https://doi.org/10.3835/plantgenome2016.09.0089>
- 390 17. Christensen O, Lund M. Genomic relationship matrix when some animals are not
 391 genotyped. *Genet Sel Evol* 2010;42(2):1-8. <http://www.gsejournal.org/content/42/1/2>
- 392 18. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: A unified
 393 approach to utilize phenotypic, full pedigree, and genomic information for genetic
 394 evaluation of Holstein final score. *J Dairy Sci* 2010;93(2):743-752.
 395 <https://doi.org/10.3168/jds.2009-2730>
- 396 19. Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Magorokosho C.
 397 Genomic selection and prediction in plant breeding. *J Crop Improv* 2011;25(3):239-261.
 398 <https://doi.org/10.1080/15427528.2011.558767>
- 399 20. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. The deviance
 400 information criterion: 12 years on. *J R Stat Soc* 2014;76(3):485-493.
 401 <https://doi.org/10.1111/rssb.12062>
- 402 21. R Core Team. R: A language and environment for statistical computing. R Foundation
 403 for statistical computing 2021. Vienna, Austria. <https://www.R-project.org/>
- 404 22. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR
 405 statistical package. *Genetics* 2014;198(2):483-495, doi:10.1534/genetics.114.164442.

Cuadro 1. Número de animales genotipados y fenotipados para el análisis de los pesos al nacer (PN), al destete (PD) y al año (PA) de una población de ganado Suizo Europeo.

Grupo	PN	PD	PA
Genotipado	300	300	300
Genotipado y fenotipado	232	218	191
Fenotipados en QRM y GBLUP	232	218	191
Fenotipados en QRH y ssGBLUP	330	267	232

QRM=Regresión cuantil usando información de marcadores, QRH=Regresión cuantil usando información de marcadores y pedigrí, GBLUP=Mejor predictor lineal insesgado genómico, ssGBLUP=Evaluación genómica en un solo paso.

Cuadro 2. Promedios de la correlación Pearson y la desviación estándar (entre paréntesis) entre los valores fenotípicos observados (y) y valores fenotípicos predichos ($\hat{y}$), cuadrado medio del error, componentes de varianza asociados al error (σ_e^2 , σ_w^2) y criterio de información de desviación (DIC).

Modelo	Cor($y, \hat{y}$)	MSE	σ_e^2 o σ_w^2	DIC
<i>Peso al nacimiento</i>				
QRM $\theta = 0.25$	0.7521	3.9973	2.7260	513.5014
	(0.0753)	(1.6108)	(1.9762)	(531.5701)
QRM $\theta = 0.50$	0.5619	7.3249	8.6297	970.7680
	(0.1501)	(0.4561)	(0.2660)	(6.9791)
QRM $\theta = 0.75$	0.7902	3.6535	2.4268	716.4237
	(0.0766)	(0.0943)	(0.4829)	(35.7161)
GBLUP	0.7924	2.3269	3.0035	803.0675
	(0.0874)	(0.2063)	(0.5578)	(31.9814)
QRH $\theta = 0.25$	0.6713	3.5026	2.3645	872.3949
	(0.1329)	(1.2848)	(1.9670)	(432.0737)
QRH $\theta = 0.50$	0.6816	2.9988	2.7372	659.1450
	(0.1253)	(0.7769)	(1.8239)	(1079.8674)
QRH $\theta = 0.75$	0.6981	4.1405	2.8610	1077.2027
	(0.1140)	(0.6187)	(0.8666)	(60.6781)
ssGBLUP	0.7055	2.4463	3.2641	1189.4282
	(0.1191)	(0.2204)	(0.4244)	(26.5023)
<i>Peso al destete</i>				

QRM $\theta = 0.25$	0.5661	476.5293	419.4138	1550.5339
	(0.2212)	(17.4612)	(23.3216)	(13.9644)
QRM $\theta = 0.50$	0.5695	357.7328	396.8138	1576.8871
	(0.2307)	(8.9681)	(47.7433)	(21.5826)
QRM $\theta = 0.75$	0.5493	175.1298	67.9660	737.2216
	(0.2196)	(47.6181)	(82.0807)	(1150.7340)
GBLUP	0.5677	294.5807	376.7794	1583.2355
	(0.2377)	(36.6279)	(24.1379)	(16.2187)
QRH $\theta = 0.25$	0.4816	644.1278	551.5150	1962.1296
	(0.0672)	(50.8464)	(64.8091)	(20.9916)
QRH $\theta = 0.50$	0.4797	366.5940	356.9005	1537.7760
	(0.0274)	(56.8604)	(238.5303)	(903.3492)
QRH $\theta = 0.75$	0.3918	216.1753	5.9471	-706.1573
	(0.0544)	(53.2417)	(11.7834)	(2034.7757)
ssGBLUP	0.4712	303.0404	421.8316	1982.3314
	(0.0502)	(37.6933)	(55.2774)	(21.9229)
<i>Peso al año</i>				
QRM $\theta = 0.25$	0.5421	1037.6529	953.6807	1487.1104
	(0.1350)	(175.2648)	(261.8652)	(35.8873)
QRM $\theta = 0.50$	0.5341	868.3651	964.4477	1524.0511
	(0.1355)	(34.0429)	(113.1832)	(12.4648)
QRM $\theta = 0.75$	0.5115	938.8244	700.7849	1284.0829
	(0.1290)	(241.2205)	(465.2109)	(402.9787)
GBLUP	0.5330	725.7579	924.8388	1526.7596
	(0.1389)	(71.3999)	(90.0089)	(11.6346)

QRH $\theta = 0.25$	0.5306	1277.9493	1172.2877	1850.7122
	(0.1411)	(44.0948)	(108.7991)	(17.2025)
QRH $\theta = 0.50$	0.5098	894.4148	1061.3157	1883.6773
	(0.1700)	(35.3996)	(129.4702)	(15.4422)
QRH $\theta = 0.75$	0.5027	915.1871	666.8830	1706.4933
	(0.1748)	(162.7629)	(413.5046)	(209.8455)
ssGBLUP	0.4712	778.6416	1071.3096	1891.9029
	(0.0502)	(84.9871)	(128.2878)	(17.5592)

Cor($\beta, \hat{\beta}$)= correlación entre fenotipos observados y predichos, MSE=cuadrados medios del error, σ_e^2 o σ_w^2 =componentes de varianza asociados al error, DIC=criterio de información de desviación.

Cuadro 3. Promedios de la correlación Pearson y desviación estándar (entre paréntesis) entre efectos de marcador “verdaderos y estimados, señales “verdaderas” y estimadas, componentes de varianza asociados al error y valores de DIC para datos simulados con diferentes grados de asimetría y proporción de valores atípicos.

Modelo	$Cor(\beta, \hat{\beta})$	$Cor(X\beta, X\hat{\beta})$	σ_e^2 o σ_w^2	DIC
$\rho = 0.95$, 5% datos atípicos				
QR $\theta = 0.25$	0.0784	0.4963	0.6821	620.5455
	(0.0034)	(0.0336)	(0.1806)	(49.3305)
QR $\theta = 0.50$	0.0766	0.4643	0.6644	665.8219
	(0.0042)	(0.0493)	(0.0703)	(16.3032)
QR $\theta = 0.75$	0.0606	0.4269	0.1438	290.6870
	(0.0132)	(0.0386)	(0.1421)	(148.9695)
GBLUP	0.0722	0.4910	0.7375	691.6503
	(0.0064)	(0.0398)	(0.0723)	(19.9391)
$\rho = 0.95$, 10% datos atípicos				
QR $\theta = 0.25$	0.0614	0.4369	0.4683	407.6496
	(0.0183)	(0.0329)	(0.4030)	(330.6304)
QR $\theta = 0.50$	0.0728	0.4579	0.7947	706.7931
	(0.0045)	(0.0420)	(0.1063)	(20.5797)
QR $\theta = 0.75$	0.0574	0.4061	0.4482	381.4644
	(0.0092)	(0.0399)	(0.3225)	(474.7138)
GBLUP	0.0654	0.4556	0.8717	731.9104
	(0.0057)	(0.0314)	(0.0890)	(21.8563)

$\rho = 0.975$, 5% datos atípicos				
QR $\theta = 0.25$	0.0773	0.4835	0.5578	582.4254
	(0.0087)	(0.0562)	(0.2523)	(83.0548)
QR $\theta = 0.50$	0.0771	0.4689	0.6369	662.0337
	(0.0074)	(0.0515)	(0.0868)	(23.8018)
QR $\theta = 0.75$	0.0598	0.4169	0.2398	219.1691
	(0.0128)	(0.0450)	(0.2033)	(444.5060)
GBLUP	0.0703	0.4804	0.7316	692.6392
	(0.0056)	(0.0333)	(0.0831)	(24.0645)
$\rho = 0.975$, 10% datos atípicos				
QR $\theta = 0.25$	0.0731	0.4386	0.8739	677.0858
	(0.0081)	(0.0789)	(0.1077)	(23.5472)
QR $\theta = 0.50$	0.0734	0.4529	0.8154	711.2935
	(0.0078)	(0.0615)	(0.0845)	(14.9809)
QR $\theta = 0.75$	0.0541	0.3945	0.3628	385.6030
	(0.0056)	(0.0583)	(0.2572)	(393.1935)
GBLUP	0.0640	0.4491	0.8913	736.7880
	(0.0077)	(0.0517)	(0.0654)	(14.8343)
$\rho = 0.999$, 5% datos atípicos				
QR $\theta = 0.25$	0.0615	0.5286	0.1535	205.6973
	(0.0144)	(0.0271)	(0.1657)	277.5807
QR $\theta = 0.50$	0.0741	0.5514	0.4860	614.2282
	(0.0037)	(0.0167)	(0.0663)	15.7647
QR $\theta = 0.75$	0.0467	0.4855	0.0166	-271.4761

	(0.0112)	(0.0150)	(0.0192)	288.4509
GBLUP	0.0737	0.5428	0.5305	625.9703
	(0.0030)	(0.0121)	(0.0353)	11.3632
$\rho = 0.999$, 10% datos atípicos				
QR $\theta = 0.25$	0.0768	0.4807	0.7817	650.8593
	(0.0080)	(0.0687)	(0.0888)	22.8417
QR $\theta = 0.50$	0.0696	0.4630	0.6154	511.5287
	(0.0148)	(0.0600)	(0.3369)	412.6645
QR $\theta = 0.75$	0.0507	0.3967	0.0204	-160.1660
	(0.0031)	(0.0505)	(0.0127)	213.0462
GBLUP	0.0659	0.4649	0.7876	709.7240
	(0.0065)	(0.0418)	(0.0528)	14.8566

$\text{Cor}(\beta, \hat{\beta})$ = correlación entre efectos de marcador “verdaderos” y estimados,

$\text{Cor}(X\beta, X\hat{\beta})$ = correlación entre señales “verdaderas” y estimadas, σ_e^2 o

σ_w^2 =componentes de varianza asociados al error, DIC=criterio de información

de desviación.

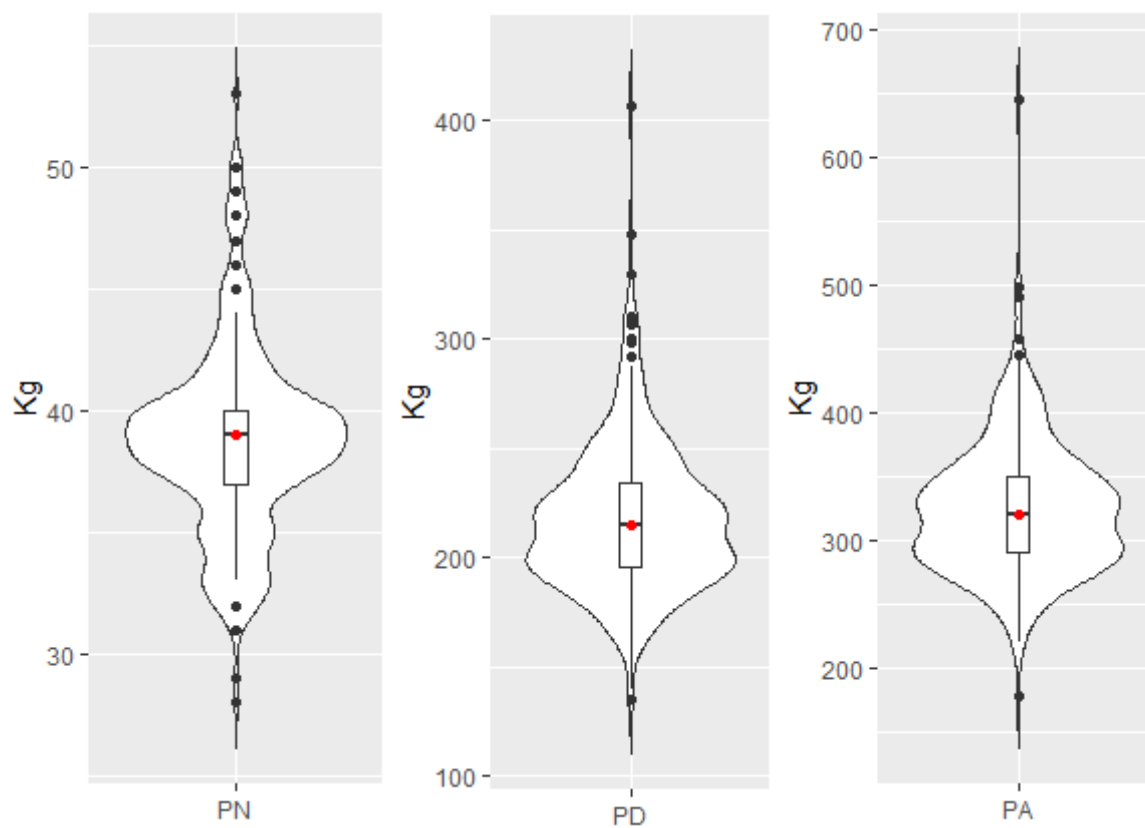


Figura 1. Gráficas de violín de peso al nacimiento (PN), al destete (PD) y al año (PA) en una población de bovinos Suizo Europeo. La media está representada por el punto rojo y la mediana por la línea horizontal dentro del recuadro.

5. CONCLUSIONES GENERALES

En este trabajo se estudió la factibilidad de incluir información genómica en evaluaciones genéticas de bovinos Suizo Europeo, utilizando diferentes metodologías estadísticas, con el fin de mejorar la tasa de progreso genético.

Una de las restricciones para incluir información genómica en las evaluaciones genéticas es la limitada disponibilidad de esta. La metodología estadística de evaluación genómica en un solo paso (ssGBLUP) es una alternativa viable para evaluar simultáneamente animales con y sin genotipo; además mejora la exactitud de predicción de los valores genéticos de los animales. La magnitud de la mejora en exactitud es mínima, pero se espera que conforme se aumente el número de animales genotipados la exactitud de predicción sea más mayor.

Otra problemática en evaluaciones tanto tradicionales como genómicas es que en los modelos estadísticos se asume normalidad, lo cual no siempre se cumple, también registros fenotípicos extremos se consideran como datos “atípicos” y son eliminados. La regresión cuantil (QR) resultó ser una alternativa viable para tratar con distribuciones no normales y colas pesadas (datos atípicos), mejora la exactitud de predicción y se puede probar diferentes cuantiles que permiten observar una mejor relación entre la variable de respuesta y las variables explicativas. Este método presentó igual o mayor poder predictivo que ssGBLUP y GBLUP y las ventajas de QR fueron más notorias con distribución asimétrica de los fenotipos y con mayor proporción de datos atípicos.

La raza de bovinos Suizo Europeo presenta potencial para mejoramiento genético; sin embargo, dado el escaso impacto del plan de mejoramiento genético de los últimos años se requiere del diseño adecuado de un programa de mejoramiento genético que incluya información genómica.

6. APÉNDICES

Apéndice 1. Código R para ajustar modelos de regresión cuantil.

```
1  rm(list=ls())
2
3  library(BGLR)
4
5  ## Warning: package 'BGLR' was built under R version 3.6.2
6
7  library(GIGrv)
8
9  setwd("C:/Users/jevvh_/Dropbox/Quantile regression/Data analysis/Quantile"
10 )
11
12 #Replace with dyn.load("util_sample_QR.dll") in Windows
13 dyn.load("util_sample_QR.dll")
14
15 #Bayesian Regularized Quantile Regression
16 #Supporting script
17
18 #Check Loss function
19 check_loss=function(t,theta)
20 {
21     ifelse(t<0,(theta-1)*t,theta*t)
22 }
23
24 #Log Likelihood
25 log_L=function(error,theta,tau)
26 {
27     n=length(error)
28     n*log(theta)+n*log(1-theta)+n*log(tau)-tau*sum(check_loss(error,theta
29 ))
30 }
31
32 #Main function
33 #y,phenotypes
34 #X, matrix for markers
35 #Z, matrix
36 #A, matrix
37 #intercept, logical, intercept is included by default,
38 #theta: quantile
39 #R2, proportion of variance explained by residuals
40 #nIter, burnIn, thin, parameters for Gibbs sampler
41 #a,b, hyperparameters for prior distributions.
42 #saveFiles, logical if true will save MCMC outputs in text files
43 #verbose, logical, if true, will print progress history
```

```

42 #When finish running it will return a list with relevant output.
43
44 BRQR=function(y,intercept=TRUE,ETA,theta=0.50,
45               R2=0.50,
46               nIter=1000,burnIn=500,thin=10,
47               a=0.01,
48               b=0.01,
49               saveFiles=TRUE,
50               verbose=TRUE)
51 {
52
53   n=length(y)
54
55   e1=(1-2*theta)/(theta*(1-theta))
56   e2=sqrt(2/(theta*(1-theta)))
57
58   vy=var(y,na.rm=TRUE)
59
60   #Observed and missing values
61   whichObs=which(!is.na(y))
62
63   missing=any(is.na(y))
64
65   if(missing)
66   {
67     whichNa=which(is.na(y))
68     nNa=length(whichNa)
69     m=mean(y,na.rm=TRUE)
70     y[whichNa]=m
71     ymiss=rep(m,nNa)
72   }
73
74   yHat=matrix(NA,nrow=nIter,ncol=n)
75
76   #Assuming all regression coefficients are set to 0
77   #See initialization below
78   #error=y DOES NOT WORK, the vectors share the same memory address!!!
79   error<-rep(NA,n)
80   for(i in 1:n)
81   {
82     error[i]=y[i]
83   }
84
85   #Initialization of terms
86
87   nLT=length(ETA)
88   fraction=R2/nLT
89
90   for(j in 1:nLT)

```

```

91 {
92   cat("Initializing linear term ",j,"\n")
93   ETA[[j]]$X2=ETA[[j]]$X^2
94
95   if(ETA[[j]]$model=="BRR")
96   {
97     ETA[[j]]$p=ncol(ETA[[j]]$X)
98     ETA[[j]]$x2=colSums(ETA[[j]]$X2)
99     ETA[[j]]$sx=colSums(ETA[[j]]$X)
100    ETA[[j]]$sumMeanXSq = sum((apply(ETA[[j]]$X,2L,mean))^2)
101    ETA[[j]]$MSx=sum(ETA[[j]]$x2)/n-ETA[[j]]$sumMeanXSq
102    ETA[[j]]$df=5
103    ETA[[j]]$Sb=fraction*vy*(ETA[[j]]$df+2)/ETA[[j]]$MSx
104    cat("df=",ETA[[j]]$df,"\n")
105    cat("Scale=",ETA[[j]]$Sb,"\n")
106    ETA[[j]]$varB=0.10
107    ETA[[j]]$beta=rep(0,ETA[[j]]$p)
108    ETA[[j]]$Beta=matrix(NA,nrow=nIter,ncol=ETA[[j]]$p)
109    ETA[[j]]$vVarB=rep(NA,nIter)
110  }else if(ETA[[j]]$model=="FIXED"){
111
112    ETA[[j]]$p=ncol(ETA[[j]]$X)
113    ETA[[j]]$beta=rep(0,ETA[[j]]$p)
114    ETA[[j]]$Beta=matrix(NA,nrow=nIter,ncol=ETA[[j]]$p)
115    ETA[[j]]$varB=1E3
116
117  }else{
118    stop("model not implemented")
119  }
120 }
121
122 if(intercept)
123 {
124   Mu=rep(NA,nIter)
125   mu=mean(y,na.rm=TRUE)
126   error=error-mu
127 }else{
128   mu=0
129 }
130
131 #Initial values
132 Tau=rep(NA,nIter)
133 tau=1
134
135 v=rexp(n=n)
136
137 #Gibbs sampler
138
139 post_logLik=rep(NA,nIter)

```

```

140
141   for(iter in 1:nIter)
142   {
143       if(verbose)
144       {
145           cat("*****")
146 *\n")
147           cat("iter=",iter,"\n")
148       }
149
150       if(intercept)
151       {
152           ytilde=error-e1*v+mu
153           mean=sum(ytilde/v)/sum(1/v)
154           variance=e2^2/(tau*sum(1/v))
155           mu_new=rnorm(1,mean=mean,sd=sqrt(variance))
156
157           #Update error
158           error=error+(mu-mu_new)
159           mu=mu_new
160
161           Mu[iter]=mu
162
163           if(verbose) cat("mu=",mu,"\n")
164       }
165
166       #Update regression coefficients and variances
167
168       for(j in 1:nLT)
169       {
170           c1=(e2^2/tau)/ETA[[j]]$varB
171           ytilde=rep(0,n)
172           out=.Call("sample_beta_c",
173                   n,
174                   ETA[[j]]$p,
175                   ETA[[j]]$X,
176                   ETA[[j]]$X2,
177                   ytilde,v,error,
178                   ETA[[j]]$beta,
179                   e1,
180                   e2,
181                   c1,
182                   tau)
183           ETA[[j]]$beta=out[[1]]
184           error=out[[2]]
185           ETA[[j]]$Beta[iter,]=ETA[[j]]$beta
186
187           if(ETA[[j]]$model=="BRR")
188           {

```

```

189         #Update varB
190         ETA[[j]]$varB=(ETA[[j]]$Sb+sum(ETA[[j]]$beta^2))/rchisq(n
191 =1,df=ETA[[j]]$df+ETA[[j]]$p)
192         ETA[[j]]$vVarB[iter]=ETA[[j]]$varB
193         if(verbose) cat("varB_",j,"=",ETA[[j]]$varB,"\n",sep="")
194     }
195 }
196
197 #Update v
198 #NOTE: DO NOT TRY TO VECTORIZE IT, IT DOES NOT WORK, READ THE
199 #PACKAGE DOCUMENTATION
200 for(i in 1:n)
201 {
202     v[i]=rgig(n=1,lambda=1/2,chi=tau*(error[i]/e2)^2,psi=tau*(e1/
203 e2)^2+2*tau)
204 }
205
206 #Update tau
207 tmp3=b+sum((error-e1*v)^2/(2*e2^2*v)+v)
208 tau=rgamma(n=1,shape=a+3*n/2,rate=tmp3)
209 Tau[iter]=tau
210 if(verbose) cat("tau=",tau,"\n")
211
212 #update yHat
213 if(missing)
214 {
215     lp=ymiss-error[whichNa]
216     ymiss=lp+e1*v[whichNa]+rnorm(n=nNa,mean=0,sd=e2*sqrt(v[whichN
217 a]/tau))
218     yHat[iter,whichNa]=ymiss
219     error[whichNa]=ymiss-lp
220 }
221
222 lp=y[whichObs]-error[whichObs]
223 yHat[iter,whichObs]=lp
224
225 #Likelihood
226 post_logLik[iter]=log_L(error[whichObs],theta,tau)
227
228 }
229
230 out=list()
231
232 index=seq(from=burnIn+thin,to=nIter,by=thin)
233
234 #For computing LogLikAtPostMean
235 error=y
236
237 for(j in 1:nLT)

```

```

238 {
239   if(saveFiles)
240   {
241     if(ETA[[j]]$model=="BRR")
242     {
243       write.table(ETA[[j]]$vVarB,file=paste("ETA_",j,"_varB.dat
244 ",sep=""),row.names=FALSE,col.names=FALSE)
245     }
246   }
247
248   ETA[[j]]$Beta=ETA[[j]]$Beta[index,]
249   ETA[[j]]$beta=colMeans(ETA[[j]]$Beta)
250   ETA[[j]]$SD.beta=as.vector(apply(ETA[[j]]$Beta,2,sd))
251
252   if(ETA[[j]]$model=="BRR")
253   {
254     ETA[[j]]$vVarB=ETA[[j]]$vVarB[index]
255     ETA[[j]]$varB=mean(ETA[[j]]$vVarB)
256     ETA[[j]]$SD.varB=sd(ETA[[j]]$vVarB)
257   }
258
259   error=error-as.vector(ETA[[j]]$X%%ETA[[j]]$beta)
260 }
261
262 if(intercept)
263 {
264   if(saveFiles)
265   {
266     write.table(Mu,file="intercept.dat",row.names=FALSE,col.names
267 =FALSE)
268   }
269
270   Mu=Mu[index]
271   out$mu=mean(Mu)
272   out$SD.mu=sd(Mu)
273   error=error-out$mu
274 }
275
276 yHat=yHat[index,]
277
278 SD.yHat=as.vector(apply(yHat,2,sd))
279 yHat=as.vector(colMeans(yHat))
280
281
282 out$yHat=yHat
283 out$SD.yHat=SD.yHat
284
285 if(saveFiles)
286 {

```

```

287     write.table(Tau,file="tau.dat",row.names=FALSE,col.names=FALSE)
288 }
289
290 Tau=Tau[index]
291 out$tau=mean(Tau)
292 out$SD.tau=sd(Tau)
293
294 #LogLikAtPostMean
295 logLikAtPostMean=log_L(error[whichObs],theta,out$tau)
296
297 post_logLik=mean(post_logLik[index])
298
299 pD = -2 * (post_logLik - logLikAtPostMean)
300 DIC = pD - 2 * post_logLik
301 DICs = -2*logLikAtPostMean+3*pD
302
303 fit=list()
304 fit$logLikAtPostMean=logLikAtPostMean
305 fit$postMeanLogLik=post_logLik
306 fit$pD=pD
307 fit$DIC=DIC
308 fit$DICs=DICs
309
310 out$fit=fit
311 out$ETA=ETA
312
313 return(out)
314
315 }

```

Apéndice 2. Código de muestreador de Gibbs en C para ajustar modelos de regresión cuantil.

```
1  /*
2
3  Compilation in Windows
4
5  R CMD SHLIB util_sample_QR.c -lRlapack -lRblas
6
7  Compilation in macOS/Linux
8
9  R CMD SHLIB util_sample_QR.c
10
11 */
12
13 #include <R.h>
14 #include <Rmath.h>
15 #include <Rinternals.h>
16 #include <Rdefines.h>
17 #include <Rconfig.h>
18 #include <R_ext/Lapack.h>
19
20 SEXP sample_beta_c(SEXP n, SEXP p, SEXP X, SEXP X2, SEXP ytilde,
21                   SEXP v, SEXP error, SEXP beta,
22                   SEXP e1, SEXP e2,
23                   SEXP c1, SEXP tau)
24 {
25     double *xj;
26     double *xj2;
27
28     double *pX;
29     double *pX2;
30     double *pb;
31     double *pe;
32     double *pytilde;
33     double *pv;
34
35     double e1_;
36     double e2_;
37     double c1_;
38     double g;
39     double tau_;
40     double b;
41     double m;
```

```

42
43     int inc=1;
44     int i,j,rows, cols;
45
46     SEXP list;
47
48     GetRNGstate();
49
50     rows=INTEGER_VALUE(n);
51     cols=INTEGER_VALUE(p);
52
53     PROTECT(X=AS_NUMERIC(X));
54     pX=NUMERIC_POINTER(X);
55
56     PROTECT(X2=AS_NUMERIC(X2));
57     pX2=NUMERIC_POINTER(X2);
58
59     PROTECT(ytilde=AS_NUMERIC(ytilde));
60     pytilde=NUMERIC_POINTER(ytilde);
61
62     PROTECT(v=AS_NUMERIC(v));
63     pv=NUMERIC_POINTER(v);
64
65     PROTECT(error=AS_NUMERIC(error));
66     pe=NUMERIC_POINTER(error);
67
68     PROTECT(beta=AS_NUMERIC(beta));
69     pb=NUMERIC_POINTER(beta);
70
71     e1_=NUMERIC_VALUE(e1);
72     e2_=NUMERIC_VALUE(e2);
73     c1_=NUMERIC_VALUE(c1);
74     tau_=NUMERIC_VALUE(tau);
75
76     for(j=0; j<cols;j++)
77     {
78         /*Column j for X and X2*/
79         xj=pX+j*rows;
80         xj2=pX2+j*rows;
81
82         /*entry j for vector beta*/
83         b=pb[j];
84
85         /*Compute ytilde*/
86         for(i=0; i<rows;i++)

```

```

87     {
88     pytilde[i]=pe[i]+xj[i]*b-e1_*pv[i];
89     }
90
91     /*Compute g*/
92     g=0;
93     for(i=0;i<rows;i++)
94     {
95     g+=xj2[i]/pv[i];
96     }
97     g+=c1_;
98
99     /*sample beta*/
100    m=0;
101    for(i=0;i<rows;i++)
102    {
103    m+=xj[i]*pytilde[i]/pv[i];
104    }
105    m=m/g;
106
107    pb[j]= m+e2_/sqrt(tau_*g)*norm_rand();
108
109    b-=pb[j];
110
111    F77_NAME(daxpy)(&rows, &b,xj,&inc, pe,&inc);
112
113    }
114
115    // Creating a list with 2 vector elements:
116    PROTECT(list = allocVector(VECSXP, 2));
117    // attaching b vector to list:
118    SET_VECTOR_ELT(list, 0, beta);
119    // attaching e vector to list:
120    SET_VECTOR_ELT(list, 1, error);
121
122    PutRNGstate();
123
124    UNPROTECT(7);
125
126    return(list);
127 }

```