



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Predicción genómica para consumo residual de alimento en toros Hereford, estudio de modelos alternativos y su habilidad predictiva

María Isabel Pravia Nin

Doctora en Ciencias Agrarias
Opción Ciencia Animal

Abril, 2025

**Predicción genómica para consumo
residual de alimento en toros Hereford,
estudio de modelos alternativos y su
habilidad predictiva**

María Isabel Pravia Nin

Doctora en Ciencias Agrarias
Opción Ciencia Animal

Abril, 2025

Tesis aprobada por el tribunal integrado por el Ing. Agr. (PhD) Fernando Baldi (presidente), el Ing. Agr. (PhD) Sebastián Munilla y la Ing. Agr. (PhD) Ana Espasandín el 23 de abril de 2025. Autora: Ing. Agr. (Mag.) María Isabel Pravia Nin. Directora: Ing. Agr. (PhD) Olga Ravagnolo. Codirector: Ing. Agr. (PhD) Ignacio Aguilar.

Agradecimientos

Agradezco a mi familia por acompañarme durante este proceso, darme el apoyo necesario para seguir adelante y por ser mi mayor fuente de motivación a lo largo de este proyecto de vida.

A mis tutores, Olga Ravagnolo e Ignacio Aguilar, por su valiosa orientación, consejos y paciencia durante el desarrollo de esta tesis, así como por sus contribuciones fundamentales en mi formación académica y profesional como investigadora en estos años de carrera.

A Elly Navajas, por sus aportes en cada uno de los artículos publicados, así como por su precisión en las correcciones de la redacción de esta tesis.

Agradezco también a mis compañeros de trabajo, quienes, a través de diversos intercambios en seminarios y discusiones, contribuyeron a la mejora de este trabajo.

Finalmente agradezco al INIA por darme la oportunidad de trabajar y formarme durante estos años, así como a las instituciones que estuvieron involucradas en proporcionar los recursos y datos necesarios para llevar a cabo esta investigación.

Tabla de contenido

	Página
<u>Página de aprobación</u>	<u>III</u>
<u>Agradecimientos</u>	<u>IV</u>
<u>Resumen</u>	<u>IX</u>
<u>Summary</u>	<u>X</u>
<u>1. Introducción</u>	<u>1</u>
1.1. Hipótesis.....	5
1.2. Objetivo general de la tesis.....	6
1.3. Objetivos específicos	6
<u>2. Revisión bibliográfica</u>	<u>8</u>
2.1. Mejoramiento genético en eficiencia de conversión.....	8
2.1.1. Antecedentes	8
2.1.2. Consumo residual de alimento	8
2.1.3. Modelación y estimación de valores de cría de consumo y RFI	10
2.1.4. Heredabilidad y correlaciones genéticas de consumo y RFI con otras características.....	14
2.2. Selección genómica.....	15
2.2.1. Introducción	15
2.2.2. Definición.....	16
2.2.3. Modelo infinitesimal	18
2.2.4. Single step genomic BLUP (SSGBLUP)	21
2.3. Predicción de valores de cría genómicos y estudio de habilidad predictiva.....	22
2.3.1. Antecedentes	22

2.3.2. Validaciones cruzadas	25
2.3.2. Método Legarra y Reverter	27
2.4. Aspectos que afectan las precisiones de las predicciones genómicas.....	29
2.4.1. Tamaño y grado de conexión entre la población de entrenamiento y candidatos a la selección.....	29
2.4.2. Heredabilidad de la característica.....	31
2.4.3. Panel utilizado de SNP.....	31
2.4.4. Uso de características indicadoras en modelos multicarácter como estrategia de mejora de las predicciones.....	31
 <u>3. Evaluation of feed efficiency traits in different Hereford populations and their effect on variance component estimation</u>	<u>34</u>
3.1. Resumen.....	34
3.2. Abstract.....	36
3.3. Introduction.....	36
3.4. Materials and methods	37
3.4.1. Animals and phenotypes.....	37
3.4.2. Subsets of data.....	37
3.4.3. Phenotypic regression models for DMI.....	38
3.4.4. Pedigree data.....	38
3.4.5. Variance component estimation.....	38
3.4.6. Univariate models	38
3.4.7 Bivariate and multivariate models.....	38
3.5. Results and discussion.....	38
3.5.1. Data description	38
3.5.2. Phenotypic regression models for DMI.....	39
3.5.3. Additive genetic variances and heritability estimates.....	39
3.5.3.1. DMI.....	39
3.5.3.2. ADG, MWT and BFAT.....	40

3.5.3.3.RFIp.....	40
3.5.4. Genetic and phenotypic correlations.....	41
3.6. Implications.....	43
3.7. Conclusions	43
3.8. References.....	43
 <u>4. Alternative models to predict residual feed intake in Hereford breed and effects on their breeding value accuracy</u>	 <u>45</u>
4.1. Resumen.....	45
4.2. Absract.....	45
4.3. Introduction.....	45
4.4. Materials and methods.....	45
4.4.1. Animals and data.....	45
4.4.2. Methods.....	47
4.4.3. Accuracy.....	48
4.4.4. Bias and re-ranking	48
4.5. Results and discussion.....	48
 <u>5. Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake.</u>	 <u>49</u>
5.1. Resumen.....	49
5.2. Abstract.....	51
5.3. Introduction	52
5.4. Materials and Methods.....	53
5.4.1. Animals and data	53
5.4.2. Models	53
5.4.2.1. Univariate model (Model 1)...	53
5.4.2.2. Multi Trait model (Model 2) ...	53
5.4.3. Validations strategies	54
5.4.4. Prediction ability statics	54

5.5. Results.....	55
5.5.1. Univariate model (Model 1)...	55
5.5.2. Multivariate model (Model 2)...	56
5.6. Discussion ..	57
5.7. Conclusions.....	60
5.8. Funding information.....	60
5.9. Conflict of interest statement	60
5.10. Data availability statement	60
5.11. Orcid	60
5.12. References..	60
<u>6. Discusión general</u>	<u>62</u>
6.1. Estimación de parámetros genéticos para eficiencia de conversión....	
.....	63
6.2. Evaluación de modelos alternativos para predecir valores de cría para RFI	65
6.3. Estudio de la habilidad predictiva de los modelos genómicos propuestos	69
<u>7. Conclusiones</u>	<u>75</u>
<u>8. Implicancias</u>	<u>7977</u>
<u>9. Bibliografía</u>	<u>79</u>
<u>10. Anexos</u>	<u>88</u>
10.1 Incorporación de información genómica en la estimación del mérito genético de consumo residual de alimento.....	88

Resumen

La eficiencia de conversión de alimento es de gran relevancia económica en los sistemas de producción de carne. El consumo residual de alimento (RFI) es el indicador utilizado en los programas de mejora genética animal. Su uso masivo ha sido restringido por los altos costos para medir en condiciones comerciales. La selección genómica es una herramienta con gran potencial para estos caracteres, ya que permite predecir el mérito genético de individuos que no pueden ser medidos. La habilidad predictiva es clave para ampliar el uso de esta herramienta. El objetivo de este trabajo fue estimar parámetros genéticos para consumo (CS) y RFI en la población de Hereford nacional, proponer modelos alternativos para RFI y estudiar la habilidad predictiva de los modelos genómicos a través de validaciones cruzadas utilizando la metodología de LR. Los resultados demostraron que tanto CS, como RFI presentan variabilidad genética y heredabilidad alta y moderada (0,45 y 0,25, respectivamente). El CS estuvo genéticamente correlacionado con el peso metabólico (PM), ganancia diaria (GD) y espesor de grasa (EG), mientras que el RFI fue genéticamente independiente de estas. Estos resultados indican una ventaja en el uso del RFI como criterio de selección, ya que la selección por menores RFI se reflejaría en progenies con menor CS sin afectar el desempeño productivo en otros caracteres. Se propusieron dos modelos alternativos para la predicción de valores de cría (VC) para RFI, el primero a partir de la combinación de los DEP obtenidos por un modelo multivariado que considera las características regresoras (CS, PM, GD, EG) y el segundo incorporando el peso al destete (PD) por ser una característica correlacionada con el CS y disponible en gran número de animales. Los DEP obtenidos por estos modelos fueron similares. Estudios de habilidad predictiva confirmaron que los modelos genómicos tuvieron una buena capacidad predictiva para estimar VC para RFI para individuos a partir de su genotipo. El modelo multivariado que incluyó el PD no demostró mejoras en la habilidad predictiva como se esperaba, y obtuvo peores indicadores respecto al modelo univariado convencional.

Palabras clave: consumo residual de alimento, heredabilidad, correlaciones genéticas, predicción genómica, habilidad predictiva

**Genomic prediction for residual feed intake in Hereford bulls: study of
alternative models and their prediction ability**

Summary

Feed efficiency is a relevant trait for animal production systems. Residual feed intake (RFI) is the indicator used in genetic improvement programs. Its widespread use has been limited by high measurement costs of measuring under commercial conditions. Genomic selection is a tool with great potential for these traits, as it allows the prediction of genetic merit in individuals that cannot be measured. The predictive ability of models is crucial for expanding the use of this tool. The main objective of this study was to estimate genetic parameters for dry matter intake (DMI) and RFI in the National Hereford population, propose alternative models for RFI and assess the predictive ability of the genomic models through cross-validations using the LR methodology. Results showed genetic variability and moderate to high heritability for both DMI (0.45) and RFI (0.25). DMI was genetically correlated with metabolic weight (MW), average daily gain (ADG) and back fat thickness (BFAT), while RFI was independent of these traits. These suggest an advantage in using RFI over DMI as a selection criterion, as selecting individuals with lower RFI will result in offspring with lower DMI without a negative impact on performance traits. Two alternative models for predicting RFI breeding values were proposed: the first, based on the combination of the DEPs obtained by a multivariate model that considers the regressor traits (DMI, MW ADG, BFAT), and a second model incorporating weaning weight (WW), as it is highly correlated with DMI and available in a large number of animals. Results showed similar DEPs between models. Predictive studies confirmed that the proposed genomic models had good predictive ability for estimating breeding values for RFI based on genotype. The multivariate model with WW showed no improvement in predictive ability, yielding worse accuracy, greater dispersion and lower precision than the univariate model.

Keywords: residual feed intake, heritability, genetic correlations, genomic prediction, predictive ability

1. Introducción

El consumo de alimento tiene alta incidencia en los costos de los sistemas de producción de carne vacuna. Dada la relación entre consumo de alimento y desempeño productivo, el incremento de los márgenes económicos está asociado a mejorar la eficiencia de conversión del alimento en producto, de manera de minimizar los niveles de consumo sin perder de vista incrementos de la producción.

En el ámbito internacional, se ha discutido bastante sobre cuál sería la característica vinculada a la eficiencia de conversión más apropiada para incluir en los programas de mejora genética animal (Nielson et al., 2013). Una primera característica para considerar es el consumo de alimento, que presenta variabilidad genética y heredabilidad moderada a alta. Sin embargo, está altamente correlacionada con las características productivas, por lo cual algunos autores sostienen que el consumo *per se* provee poca información sobre la eficiencia de conversión, ya que la selección por menor consumo puede traer cambios no deseados en otras características de importancia económica (Beef Improvement Federation, 2010). En este sentido, la consideración del consumo en conjunto con otras características fue propuesta como una alternativa recomendable.

Si bien existen diferentes formas de medir la eficiencia de conversión, el consumo de alimento residual (RFI, por su sigla en inglés, *residual feed intake*) ha sido priorizado. El RFI fue definido por Koch et al. (1963) como la diferencia entre el consumo real y el consumo predicho basado en los requerimientos de mantenimiento y producción. Este indicador ha sido la medida de eficiencia más utilizada e incorporada por los programas de mejora en varios países y razas, ya que permite reducir el consumo de alimento, sin afectar el desempeño productivo.

Las posibilidades de mejora genética para eficiencia de conversión, ya sea por RFI u otro indicador, vienen asociadas al uso de la selección genómica. Esta herramienta de alto potencial para incrementar la tasa de mejora genética además se presenta como una gran oportunidad para la incorporación en programas de mejora genética de este tipo de características de difícil y costosa medición. En los últimos años, el uso de la selección genómica ha crecido exponencialmente en bovinos para

leche, así como en bovinos de carne, lo que ha permitido acelerar el progreso genético principalmente en características que no son medidas en el individuo candidato a la selección, como la producción de leche, la terneza de la carne o el consumo de alimento y eficiencia de conversión (Goddard, 2012).

La selección genómica se basa en el uso de méritos genéticos obtenidos a partir de la información de un conjunto de miles de marcadores moleculares distribuidos a lo largo del genoma. Los marcadores utilizados son los polimorfismos de nucleótido simple (SNP), que son variaciones a nivel del genoma, las cuales son usadas para predecir el potencial genético de un animal con base en su efecto conjunto sobre los caracteres de interés (Meuwissen et al., 2001). Un aspecto clave en la implementación de la selección genómica es el diseño de las poblaciones de entrenamiento o referencia. Estas son poblaciones de animales que cuentan con información fenotípica para la o las características de interés y genotipo registrado. A partir de estas poblaciones es posible predecir el mérito genético de animales que no cuenten con fenotipo propio o de sus descendientes. En la construcción de la población de referencia o entrenamiento, el parentesco entre los animales que la componen y su variabilidad fenotípica tienen un rol fundamental, ya que van a tener un impacto directo en la precisión de las predicciones de los méritos genéticos (Calus et al., 2013). A nivel internacional, han existido iniciativas para compartir fenotipos y genotipos y generar poblaciones de entrenamiento robustas para características de difícil medición. Sin embargo, el conocimiento de cómo se manifiesta la característica en cada país es un aspecto importante a tener en cuenta para la unión de información de distintos orígenes.

Los programas de mejoramiento genético en las principales razas de bovinos de carne en nuestro país cuentan con varios años de implementación y han demostrado ser efectivos. Las tendencias genéticas para las características vinculadas al crecimiento muestran un sostenido progreso genético, así como aquellas vinculadas a las características carniceras (www.geneticabovina.com.uy). Por otro lado, la incorporación de la eficiencia de conversión y del consumo de alimento ha sido muy limitada, a pesar de la alta relevancia económica, en Uruguay e internacionalmente.

La reciente incorporación de la eficiencia de conversión en las evaluaciones genéticas bovinas es debida principalmente a las dificultades de medición del consumo de alimento en forma individual en las cabañas de forma directa, a diferencia de otros criterios de selección como los pesos y mediciones de composición corporal *in vivo* por ultrasonido. En Uruguay, las mediciones en eficiencia de conversión se iniciaron en 2014 en la raza Hereford a partir de la instalación de comederos automáticos que permiten la medición del consumo individual. Esta iniciativa fue diseñada con el objetivo de construir la población de referencia para aplicar selección genómica en esta característica. Dadas las exigencias técnicas de medición, solo una baja proporción de los candidatos a la selección son evaluados genéticamente a través de su fenotipo, mientras que la selección genómica viabiliza que otra proporción de individuos que son genotipados puedan disponer de estimaciones de méritos genéticos a partir de predicciones genómicas a través de sus genotipos.

Al contar con las instalaciones para medir el consumo de alimento individual, se establecieron los protocolos para su registro y desde entonces se realizan pruebas de eficiencia de conversión en las cuales se ha medido el consumo individual de alimento en toros y novillos Hereford (Pravia et al., 2015). Al momento de comenzar con esta tesis, la población de referencia de animales medidos en el país era de más de 780 animales con fenotipo para consumo de alimento y sus genotipos, lo que permitió plantearse los estudios que se desarrollarán en este trabajo.

Si bien estudios internacionales indicaban la existencia de variabilidad genética en RFI en bovinos y su potencial de mejora a través de la selección, es relevante contar con estimaciones de parámetros genéticos propios para poder incluirlo en los programas de mejora genética nacionales. Los primeros estudios nacionales permitieron comprobar la existencia de variabilidad genética para RFI (Ravagnolo et al., 2018). Al día de hoy, al contar con un mayor volumen de información recolectada, es posible profundizar en el conocimiento del RFI, estimar la variabilidad genética y heredabilidad, así como los distintos rasgos asociados al consumo y las correlaciones genéticas con otras características de interés. Esto permitirá conocer el

potencial por selección directa, así como las respuestas en otras características correlacionadas, y podrá facilitar la inclusión del consumo en índices de selección.

Los resultados generados en este trabajo contribuyen a comprender y evaluar las ventajas o desventajas de considerar el consumo de alimento o el RFI como la característica por incorporar en los programas de mejora nacionales. También será posible determinar si la conjunción de distintas fuentes de información disponible (datos de Uruguay y Canadá) para incrementar la población de referencia es posible, a partir del análisis de las relaciones entre consumo y las características predictoras, y si esto tiene implicancias sobre las predicciones de los valores de cría.

La definición de los modelos genéticos más adecuados para consumo y RFI es otro aspecto fundamental para estudiar, a los efectos de estimar el mérito genético de los animales de forma precisa y confiable. Diferentes alternativas de modelación y evaluación genética para RFI son posibles y tienen implicancias directas sobre la habilidad predictiva y precisión de las predicciones de méritos genéticos. Al momento, la evaluación genética para el RFI (RFI fenotípico) es calculada en dos etapas: la primera es la obtención del valor fenotípico de RFI por regresión lineal en función del peso, ganancia y espesor de grasa, y posteriormente esa característica es evaluada en un modelo de predicción genómica univariado corregido por efectos fijos adicionales (Ravagnolo et al., 2018). Otra alternativa es mediante el uso de las predicciones obtenidas a partir de un modelo multivariado con base en las características registradas (consumo, peso metabólico, ganancia diaria y espesor de grasa) y luego combinadas en un índice para predecir el consumo residual. En esta alternativa se obtienen los méritos genéticos de cada característica considerada para la estimación del RFI, los cuales se relacionan en una combinación lineal para calcular un RFI *genético* (Crews et al., 2010; MacNeil et al., 2013). De acuerdo con Kennedy et al. (1993), estas dos alternativas son equivalentes. Una tercera alternativa es el uso de características predictoras o características correlacionadas con el consumo, cuya información está disponible para un mayor número de animales (MacNeil et al., 2013; Manzanilla-Pech et al., 2017). El uso de características altamente correlacionadas permitiría predecir el mérito genético para consumo o eficiencia para un mayor número de animales, al mismo tiempo eliminar el posible

sesgo causado por la selección de los animales que participan de las pruebas de consumo (Pollak et al., 1984; Pszczola et al., 2013; Lu et al., 2015). Estas alternativas son desarrolladas e investigadas en esta tesis.

La evaluación de los distintos modelos propuestos debe contemplarse en un contexto donde la predicción genómica juega un rol determinante en las respuestas a la selección para este tipo de características donde son pocos los animales que pueden contar con el fenotipo. Por ello, estudios de habilidad predictiva a través de validaciones cruzadas e indicadores son fundamentales a la hora de evaluar cuál modelo se comporta mejor. Históricamente, la habilidad predictiva de los modelos ha sido evaluada a través de validaciones cruzadas y distintas estrategias de partición de los datos (Saatchi et al., 2011). Recientemente, Legarra y Reverter (2017, 2018) han propuesto otras alternativas para evaluar la habilidad predictiva de las predicciones genómicas, en las cuales se incluye el uso de indicadores que miden la dispersión, el sesgo y precisión en la habilidad predictiva, contrastando las predicciones de méritos genéticos obtenidas de evaluaciones genéticas con datos completos y parciales. Estos indicadores son los considerados y discutidos en este trabajo.

1.1. Hipótesis

1. Diferentes subpoblaciones de animales (según países o sexo) pueden tener heterogeneidad de sus componentes de varianzas y relaciones parciales entre las variables de consumo y las variables regresoras, lo que puede afectar las estimaciones de las predicciones genéticas.

2. El consumo de alimento está correlacionado genéticamente con otras características de importancia económica de los sistemas de producción de carne como el peso vivo y la ganancia diaria, mientras que el RFI, definido como la diferencia entre el consumo observado y el consumo predicho a partir del desempeño productivo, en peso, ganancia y cobertura de grasa, es independiente de dichos caracteres.

3. El uso de la información genómica en la predicción del mérito genético para RFI de los animales sin fenotipo es una herramienta que permite predecir el valor de cría para dichos animales de forma precisa.

4. Las predicciones de los méritos genéticos para RFI obtenidos a partir de la combinación de los DEP obtenidos por un modo multivariado que considere las características que lo componen (consumo, peso metabólico, ganancia diaria y espesor de grasa dorsal) son equivalentes a las estimaciones de mérito genético obtenidas a través del modelo para RFI unicarácter.

5. La incorporación de características predictoras o indicadoras a través de modelos multivariados en la población de referencia, así como en la población por predecir, permitirá mejorar las predicciones de RFI respecto a la evaluación de estas a través de un modelo unicarácter fenotípico.

1.2. Objetivo general de la tesis

Estudiar la variabilidad genética de consumo y consumo residual de alimento (RFI) en la población de animales Hereford en Uruguay y la habilidad predictiva de diferentes modelos genómicos alternativos para la predicción del mérito genético de RFI.

1.3. Objetivos específicos

1. Estimar parámetros genéticos (varianza genética aditiva y residual, heredabilidad y correlaciones genéticas) para consumo, ganancia diaria, peso metabólico y espesor de grasa dorsal y consumo residual de alimento para las poblaciones Hereford de Uruguay, Canadá y población conjunta.

2. Proponer modelos alternativos para la predicción de RFI, entre los cuales se consideran un modelo univariado para RFI versus modelos multivariados a partir de las variables regresoras de consumo y un modelo que incluya otra característica correlacionada con consumo de alimento que sea registrada en la evaluación genética nacional.

3. Estudiar la habilidad predictiva de dos modelos genómicos propuestos para el consumo residual de alimento a través de distintos indicadores.

El desarrollo de este trabajo de tesis será presentado en varios capítulos. El capítulo 2 consiste en el planteo del marco teórico o revisión bibliográfica. Los

capítulos 3, 4 y 5 contienen los tres trabajos centrales, dos fueron publicados en revistas arbitradas y uno en un resumen expandido y arbitrado de un congreso internacional. El capítulo 6 contiene la discusión general, el capítulo 7, las conclusiones generales y el capítulo 8 unas breves implicancias.

2. Revisión bibliográfica

2.1. Mejoramiento genético en eficiencia de conversión

2.1.1. Antecedentes

Los programas de mejora genética en bovinos para carne se han enfocado en características vinculadas a los ingresos como velocidad de crecimiento, peso vivo, calidad de carne o reproducción. Poco se ha trabajado en características vinculadas con los costos, como el consumo de alimento o la sanidad, debido, en mayor medida, a la dificultad de medición. El consumo de alimento ha sido identificado como el principal costo de los sistemas de producción (Pravia et al., 2014), por lo cual en los últimos años se han realizado mayores esfuerzos por incluir esta característica en las evaluaciones genéticas, se han elaborado protocolos para su registro, así como estudios de cómo analizar e interpretar los registros, estimación de valores de cría para consumo y eficiencia y el estudio de correlaciones genéticas con otras características de importancia económica. En los últimos años, se agrega la creciente preocupación por la sostenibilidad ambiental, donde la eficiencia sobre el uso de los recursos naturales y, por ende, el menor consumo de alimento para producir los mismos kilos de producto toman también otra dimensión importante por su relación con las emisiones de metano (Basarab et al., 2013; Navajas et al., 2022).

2.1.2. Consumo residual de alimento

La consideración de consumo *per se* en forma individual provee poca información sobre la eficiencia de conversión, ya que la reducción del consumo de alimento puede traer aparejada la disminución en la producción (Crews, 2005). Frente a esto, varios investigadores plantean la inclusión del consumo en un índice de selección o, de lo contrario, de indicadores que relacionen las diferentes características (Herd et al., 2003; Nielsen et al., 2013). Distintos indicadores se han propuesto como rasgos vinculados a la eficiencia de alimentación (Berry y Crowley, 2013). El consumo residual del alimento (RFI) fue definido como la diferencia entre el consumo real observado y el esperado requerido para mantenimiento y producción

(Koch et al., 1963) y es, a diferencia de otras medidas de eficiencia, fenotípicamente independiente de las características de crecimiento. Se basa en la partición del consumo en proporciones requeridas para el estado corporal y nivel de producción (peso y ganancia) y una porción residual no explicada por las variables regresoras que se relaciona con la eficiencia metabólica real.

El modelo más simple de cálculo de RFI es el siguiente:

$$y = \beta_0 + \beta_1 (ADG) + \beta_2 (WT)^{0,75} + \text{otros efectos fijos} + e \quad [0]$$

donde y es el consumo diario en kilogramos de MS (DMI, *dry matter intake*, en su sigla en inglés); β_0 es el intercepto; β_1 , el coeficiente de regresión parcial de la ganancia diaria (ADG, por su nombre en inglés, *average daily gain*) sobre el consumo y β_2 , el coeficiente de regresión parcial del peso metabólico ($WT^{0,75}$, por su nombre en inglés, *metabolic weight*) sobre el consumo. Otros efectos fijos pueden ser considerados como prueba, corral, sexo. El residuo, “e”, se espera tenga media 0 y distribución normal.

Estudios posteriores recomendaron la inclusión del espesor de grasa dorsal al modelo (BFAT, *back fat* en su sigla en inglés), debido a que para una misma ganancia es posible encontrar diferencias en la demanda de energía para la deposición de grasa y proteína. La inclusión de BFAT en el modelo permite, de esta manera, independizarlo de esta variable (Basarab et al., 2013; Berry y Crowley, 2013).

$$y = \beta_0 + \beta_1 (ADG) + \beta_2 (WT)^{0,75} + \beta_3 BFAT + \text{otros efectos fijos} + e \quad [1]$$

El residuo “e” de los modelos 0 y 1 representa el RFI y es aquella variación no explicada en el consumo observado por los demás regresores. Este componente residual del consumo puede ser utilizado para identificar animales que se desvían de sus niveles de consumo esperados en función de su ganancia diaria, peso metabólico y nivel de engrasamiento (Crews, 2005).

Por las propiedades de la regresión lineal el residuo del modelo (“e” o RFI) es independiente fenotípicamente de los términos de regresión parciales utilizados en el modelo de estimación (ADG, MW, BFAT) llamados por algunos autores *energy sinks*. Esto ha sido verificado en términos fenotípicos (Arthur et al., 2001, Basarab et al.,

2003). Sin embargo, el RFI estimado por regresión fenotípica puede no ser genéticamente independiente de los *energy sinks* regresores del modelo, lo que podría tener implicancias en las respuestas a la selección a largo plazo (Crews, 2005). Kennedy et al. (1993) demostraron que, a pesar de que la derivación fenotípica de RFI asegura que la correlación fenotípica entre RFI y sus componentes de producción sea cero, las correlaciones genéticas entre estos pueden ser diferentes a cero, denominando a ese indicador *RFI fenotípico*. De acuerdo con estos autores, si se asume linealidad, la variabilidad en RFI está completamente determinada por la variabilidad en consumo y la regresión parcial de sus características componentes sobre el consumo. De la misma manera, la heredabilidad de RFI, así como la respuesta a la selección para esta, está determinada por los parámetros genéticos y fenotípicos de las características que la componen.

Varios autores sostienen que la combinación de consumo y crecimiento para calcular una medida de eficiencia no agrega información adicional a lo que puede ser obtenido directamente de las características que lo componen (Crews et al., 2010; Kennedy et al., 1993; Lu et al., 2015; MacNeil et al., 2011). Por esta razón sustentan que la definición de RFI con base en regresiones genéticas es la mejor manera de considerarla y es equivalente a la selección por un índice para consumo, producción y peso, restringida para mantener la producción y el peso constantes.

2.1.3. Modelación y estimación de valores de cría de consumo y RFI

La modelación más convencional para la predicción de eficiencia a través del RFI incluye dos etapas: primero, la derivación lineal de RFI a partir del consumo registrado individualmente, como fue expresado en [1]. Posteriormente, el RFI es considerado en un análisis BLUP univariado como la variable de respuesta:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [2]$$

donde $\mathbf{y}$ son las observaciones de RFI; $\boldsymbol{\beta}$, los efectos fijos del modelo; $\boldsymbol{\mu}$, la estimación del mérito genético del animal para RFI y $\mathbf{X}$ y $\mathbf{Z}$, las matrices de incidencia que relacionan las observaciones con los efectos fijos y los efectos aleatorios (efecto animal), respectivamente. Se asume que $\mathbf{u} \sim N(0, A\sigma_u^2)$, y $\mathbf{e} \sim N(0, I\sigma_e^2)$, donde A es

la matriz de parentesco basada en la genealogía. La matriz de parentesco A puede ser reemplazada por la matriz de parentesco genómico G o la matriz H que combina el parentesco basado en la genealogía con el parentesco basado en información molecular (GBLUP y ssGBLUP, respectivamente) y de esa forma, la implementación del modelo permite incorporar información genómica en la predicción de méritos genéticos.

Kennedy et al. (1993) demostraron que incluir los componentes de RFI (consumo, ganancia y peso) en un índice es matemáticamente equivalente a considerar el RFI, con la ventaja de que el consumo es un concepto mucho más sencillo de comprender que el RFI. Adicionalmente, los animales que participan de las pruebas de eficiencia son en general preseleccionados, lo que podría conducir a un sesgo de la estimación de valores de cría para esta característica (Pollak et al., 1984). Los modelos univariados para RFI no consideran los efectos de esta preselección e incluso modelos multivariados incluyendo peso y ganancia de animales en las pruebas también resultan en evaluaciones sesgadas, especialmente si consideran solo la información de los animales que participan de las pruebas. Esto también fue demostrado por Lu et al. (2015), quienes reafirman el concepto de que el RFI tiene algunas restricciones desde el punto de vista conceptual, estadístico y práctico. En primer lugar, el RFI no es una característica observable, con lo cual es difícil de explicar a los criadores. Segundo, cualquier análisis de regresión utilizado para derivar RFI implícitamente asume que todas las covariables consideradas para su cálculo (o *energy sinks*) son registradas sin error de medición o son interpoladas a través de regresiones aleatorias y, en esos casos, esos estimadores pueden ser inconsistentes y sesgados. Tempelman et al. (2015) demostraron que el análisis conjunto de poblaciones diversas puede llevar al sesgo debido a la heterogeneidad de los estimadores de los coeficientes de regresión entre consumo y los *energy sinks*, lo que puede afectar las estimaciones del RFI. Adicionalmente, si alguno de los *energy sinks* no fue registrado o es eliminado para algún animal, ninguno de los demás registros para dicho animal podrá ser utilizado para calcular el RFI. Lu et al. (2015) proponen un modelo multivariado que considere todos los animales a pesar de que puedan tener algún faltante de los registros de los *energy sinks*. Esta alternativa,

también propuesta por MacNeil et al. (2013) como RFI *genético* obtenido por un modelo multicarácter a partir de los propios caracteres regresores, genera mejoras en las precisiones, además de facilitar la incorporación de los caracteres en los índices de selección.

Bajo los mismos argumentos, MacNeil et al. (2011) propusieron el uso de modelos multicarácter en la evaluación genética de consumo para la raza Angus en Estados Unidos, donde la estimación de valor de cría de RFI se realiza a través un modelo multivariado a partir de las características consumo, peso al destete, ganancia predestete y espesor de grasa:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 \\ \mathbf{X}_3\boldsymbol{\beta}_3 \\ \mathbf{X}_4\boldsymbol{\beta}_4 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1\mathbf{u}_1 \\ \mathbf{Z}_2\mathbf{u}_2 \\ \mathbf{Z}_3\mathbf{u}_3 \\ \mathbf{Z}_4\mathbf{u}_4 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \mathbf{e}_4 \end{bmatrix} \quad [3]$$

Donde $\mathbf{y}_i$ es el vector de la $i^{\text{ésima}}$ característica: consumo, peso al destete, ganancia posdestete y espesor de grasa, respectivamente; $\mathbf{X}_i$, las matrices de incidencia que relacionan los datos con los efectos fijos ($\boldsymbol{\beta}_i$) y $\mathbf{Z}$, la matriz que relaciona los fenotipos con los efectos genético-aditivos de cada característica ($\mathbf{u}_i$, $i=1$ a 4). El error ($\mathbf{e}_i$) constituye los residuos de los efectos no explicados por el modelo. Asume que los efectos genéticos aleatorios ($\mathbf{u}_1$) tienen media 0 y varianzas:

$$\text{var} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\boldsymbol{\sigma}_{u_1}^2 & A\sigma_{\mu_1 u_2} & A\sigma_{\mu_1 u_3} & A\sigma_{\mu_1 u_4} \\ A\sigma_{\mu_2 u_1} & \mathbf{A}\boldsymbol{\sigma}_{u_2}^2 & A\sigma_{\mu_2 u_3} & A\sigma_{\mu_2 u_4} \\ A\sigma_{\mu_3 u_1} & A\sigma_{\mu_3 u_2} & \mathbf{A}\boldsymbol{\sigma}_{u_3}^2 & A\sigma_{\mu_3 u_4} \\ A\sigma_{\mu_4 u_1} & A\sigma_{\mu_4 u_2} & A\sigma_{\mu_4 u_3} & \mathbf{A}\boldsymbol{\sigma}_{u_4}^2 \end{bmatrix}$$

donde $\mathbf{A}$ representa la matriz de relaciones de parentesco. En análisis genómicos, la matriz de parentesco $\mathbf{A}$ es reemplazada por la $\mathbf{G}$ o por $\mathbf{H}$ (Aguilar et al., 2010). Los efectos residuales se asumen con media 0 y varianzas:

$$\text{var} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} \mathbf{I}\boldsymbol{\sigma}_{e_1}^2 & I\sigma_{e_1 e_2} & I\sigma_{e_1 e_3} & I\sigma_{e_1 e_4} \\ I\sigma_{e_2 e_1} & \mathbf{I}\boldsymbol{\sigma}_{e_2}^2 & I\sigma_{e_2 e_3} & I\sigma_{e_2 e_4} \\ I\sigma_{e_3 e_1} & I\sigma_{e_3 e_2} & \mathbf{I}\boldsymbol{\sigma}_{e_3}^2 & I\sigma_{e_3 e_4} \\ I\sigma_{e_4 e_1} & I\sigma_{e_4 e_2} & I\sigma_{e_4 e_3} & \mathbf{I}\boldsymbol{\sigma}_{e_4}^2 \end{bmatrix}$$

donde **I** representa una matriz de identidad con rango igual al número de animales con registro.

Luego de resolver las ecuaciones multicarácter, se obtienen las soluciones de **u** para cada característica, la cuales se combinan en el siguiente índice para obtener el mérito genético para eficiencia como:

$$RFI_g = \hat{u}_1 - b_2\hat{u}_2 - b_3\hat{u}_3 - b_4\hat{u}_4 \quad [4]$$

donde los coeficientes b son obtenidos como:

$$V^{-1}C = \begin{bmatrix} b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad [5]$$

$$V = \begin{bmatrix} \sigma_{u_2}^2 & \sigma_{\mu_2\mu_3} & \sigma_{\mu_2\mu_4} \\ \sigma_{\mu_2\mu_3} & \sigma_{u_3}^2 & \sigma_{\mu_3\mu_4} \\ \sigma_{\mu_2\mu_4} & \sigma_{\mu_3\mu_4} & \sigma_{u_4}^2 \end{bmatrix}, \quad y \quad C = \begin{bmatrix} \sigma_{\mu_1\mu_2} \\ \sigma_{\mu_1\mu_3} \\ \sigma_{\mu_1\mu_4} \end{bmatrix}$$

Con esta aproximación, se obtiene un valor de cría para cada una de las características, las cuales posteriormente se relacionan a través de una combinación lineal para calcular un RFI *genético* (Crews et al., 2010; Kennedy et al., 1993; MacNeil et al., 2011; MacNeil et al., 2013). Esta alternativa tiene la ventaja de considerar la variación genética de cada una de las características, así como la covarianza entre estas, lo que facilita la incorporación del consumo en un índice.

Incluir características indicadoras en la evaluación genética tiene la potencialidad de mejorar la precisión de la evaluación de consumo y extender la evaluación de esta característica a animales que no tienen la característica registrada (MacNeil et al., 2011; Nielsen et al., 2013). Además, las evaluaciones multicarácter permiten controlar el sesgo que se da por la preselección de los animales que participan de las pruebas de consumo y eficiencia de conversión (MacNeil et al., 2011; Pollak et al., 1984).

La incorporación de características indicadoras a través de modelos multicarácter también ha sido aplicada en predicción genómica como una manera

económica de incrementar significativamente la precisión de la característica de interés por predecir, además de remover el sesgo de la característica por ser medida en pocos animales de la población de entrenamiento (Pszczola et al., 2013).

2.1.4. Heredabilidad y correlaciones genéticas de consumo y RFI con otras características

Conocer la variabilidad genética de una característica y su heredabilidad es muy importante, ya que determina el potencial de respuesta a la selección esperable. Una recopilación de estimaciones de parámetros genéticos para eficiencia en bovinos de carne llevado a cabo por Berry y Crowley (2013) reportan que el consumo de alimento es altamente heredable y encuentran valores de h^2 de entre 0,30 y 0,70, con valores promedio de 0,40. En cuanto al RFI, las estimaciones de heredabilidad reportadas se encontraron entre 0,07 y 0,62, con un valor promedio de 0,33. Si nos concentramos en estudios en razas británicas, una investigación en la raza Angus reporta h^2 de 0,37 y 0,38 para consumo y RFI, respectivamente (Arthur et al., 2001), mientras que en la raza Hereford las estimaciones fueron de h^2 de 0,31 y 0,16 para consumo y RFI, respectivamente (Herd y Bishop, 2000). Estos valores de heredabilidad moderada indican que existe potencial de respuestas a la selección para ambas características.

Por otro lado, las correlaciones genéticas entre consumo y RFI han sido de 0,64 en ganado Hereford (Herd y Bishop, 2000) y 0,69 en Angus (Arthur et al., 2001). Con respecto a las correlaciones genéticas entre estas características y demás caracteres de interés, se ha reportado que el consumo está correlacionado genéticamente con el peso (0,65) y la ganancia (0,54), mientras que las correlaciones entre el RFI con peso y ganancia fueron de -0,06 y -0,04, respectivamente (Arthur et al., 2001).

Con respecto a las correlaciones entre el consumo y otros pesos como el destete o peso al año, estudios muestran una alta correlación genética entre estas características (correlaciones entre 0,28 y 0,56, respectivamente). Por otro lado, el RFI, a pesar de no estar correlacionado con el peso vivo y ganancia diaria en prueba,

se observó que presenta alta y moderada correlación negativa con el peso directo al destete y al año (-0,45 y -0,26, respectivamente) (Arthur et al., 2001).

Se han reportado efectos maternos sobre consumo y sobre RFI en Angus y correlaciones genéticas positivas (0,22) entre el consumo y los efectos maternos en peso al destete (Arthur et al., 2001).

En cuanto a las correlaciones genéticas con características de la canal, el consumo presenta una correlación genética positiva con espesor de grasa a nivel de la 12.^a-13.^a costilla y a nivel de cuadril moderadas (0,27 y 0,14), y con área de ojo de bife (AOB) 0,43, mientras que las correlaciones genéticas entre RFI y estas características fueron cercanas a cero para grasa en cuadril y AOB, y leves y positivas para RFI y grasa en costilla (0,17) (Arthur et al., 2001).

2.2. Selección genómica

2.2.1. Introducción

La mayoría de las características de relevancia económica están controladas por el efecto conjunto de varios genes y su interacción con el ambiente. La variación genética es debida a la variación en la secuencia de ADN entre individuos, por lo que la selección de individuos que contengan los alelos más favorables sería lo deseado. Sin embargo, la mayoría de los sitios donde se encuentran los genes que controlan las características relevantes son desconocidos (Goddard, 2012). Tradicionalmente, la mejora genética de estas características se ha basado en el uso de datos fenotípicos e información de parentesco, para estimar los efectos aditivos combinados de todos los genes para estimar valor de cría para cada animal. Bajo estos supuestos, Henderson (1975) propuso las ecuaciones de modelos mixtos y a partir de estas ecuaciones predicen las mejores estimaciones lineales insesgadas (BLUE) de efectos fijos y predictores (BLUP) de efectos aleatorios.

El desarrollo de paneles de genotipados de alta densidad a menores costos, así como el desarrollo de metodologías para la inclusión de los genotipados en las evaluaciones genéticas, han permitido un incremento en el uso de esta tecnología con

el fin de ser aplicada a la producción ganadera (Goddard, 2012). La principal motivación de incorporar información molecular en la ganadería fue la expectativa de que llevaría a mayores progresos genéticos (Meuwissen et al., 2001).

En los inicios, la identificación de algunos QTL asociados a características de interés y su aplicación en la selección asistida por marcadores no tuvo el éxito esperado. Como la mayoría de las características de relevancia económica están afectados por varios genes, el beneficio de la selección por pocos QTL es limitado por la muy baja proporción de la varianza genética explicada por determinados QTL (Meuwissen y Goddard, 1996). En la medida que fue posible genotipar miles de polimorfismos de nucleótidos simple (SNP) distribuidos a lo largo de todo el genoma, se pudo implementar la selección genómica.

2.2.2. Definición

La selección genómica (SG) es una herramienta de mejoramiento genético que permite seleccionar individuos en base a predicciones de sus valores genéticos, las cuales se obtienen a partir de su propio genotipo. Estas predicciones utilizan paneles de alta densidad de marcadores moleculares conocidos como SNP (por su nombre en inglés, *single nucleotide polymorphisms*), lo que permite estimar con mayor precisión el mérito genético de cada animal (Goddard, 2009). El concepto inicial de la selección genómica fue que los SNP o marcadores deberían estar en desequilibrio de ligamiento (LD) con los polimorfismos causales (Meuwissen et al., 2001). Sin embargo, luego se vio que, como los genes que codifican están distribuidos por todo el genoma, los métodos estadísticos utilizados no requieren de la identificación de los sitios específicos que causan variación en la característica (Goddard, 2009).

A través de las predicciones genómicas, se obtienen los valores de cría de los individuos como la sumatoria de todos los efectos de los marcadores (SNP) estimados a partir de la población de referencia con fenotipo conocido. Este método estima simultáneamente el efecto de un gran número de SNP y predice el valor de cría como la sumatoria de cada uno de los efectos del SNP sobre la característica en cuestión para obtener así los valores de cría genómicos (GEBV) (Meuwissen et al., 2001). La

asociación entre cada SNP y el desempeño son estimados a partir de las poblaciones de referencia o entrenamiento.

La población de referencia o entrenamiento consiste en poblaciones de animales con fenotipo y genotipo conocido, las que permiten cuantificar los efectos de cada SNP para luego ser aplicado en una población de animales candidatos a la selección de la cual solo se disponga de sus genotipos (Goddard, 2012).

Las precisiones de las predicciones genómicas dependen en gran medida del tamaño de la población de referencia (Goddard, 2009) y la heredabilidad de la característica, así como de la relación entre la población de referencia y los candidatos a la selección (Habier et al., 2007). Este concepto será ampliado en otro punto de la revisión.

La principal ventaja de la selección genómica sobre los métodos tradicionales es mayor donde la aplicación de los mismos es difícil de implementar. Esto es donde el fenotipo de interés no puede ser observado en los candidatos a la selección como la producción de leche, mediciones de carcasa que no pueden ser medidas en el toro candidato a la selección, sino en progenies del toro. Otra de las ventajas de la selección genómica (SG), es que permite identificar de forma precisa animales superiores a edades tempranas, las cuales se basa en estimaciones de mérito genético derivados de la información del ADN. Estas predicciones para animales jóvenes son considerablemente más precisas que las realizadas a partir del mérito genético de sus padres (Calus, Berry, et al., 2013). Este incremento en las precisiones a edades tempranas ha llevado a que la selección genómica sea una herramienta muy atractiva principalmente en los programas de mejoramiento genético en ganadero lechero en caracteres que se miden en hembras o caracteres costosos de medir. En ganado de carne, las principales ventajas de la SG sobre los métodos tradicionales se presenta también para la selección de caracteres donde el fenotipo no puede ser medido en el individuo, es muy costoso de medir o se mide muy tarde en el tiempo.

Otra ventaja de la selección genómica es que la presencia de marcadores ha permitido conocer que largos segmentos de cromosomas son compartidos entre familias cercanas, pero también que hay segmentos aislados que son idénticos en estado y son compartidos por ancestros más distantes (VanRaden et al., 2009). Con

el uso de la genómica, esta información también puede ser incorporada, lo que ha permitido detectar errores de pedigrí (Heaton et al., 2002) y establecer relaciones de parentesco genómico (Meuwissen y Goddard, 2010).

2.2.3. Modelo infinitesimal

Los primeros trabajos de predicción genómica se basaron en el modelo infinitesimal presentado por Meuwissen et al. (2001), quien fundamentó que la predicción del mérito genético de un individuo consiste en la sumatoria de pequeñas variaciones en los SNP y se describe como:

$$Y_i = \mu + \sum (SNP_{ijk}) + e$$

donde Y_i es el fenotipo considerado para cada individuo i , el cual puede ser un fenotipo corregido por algún modelo o un valor de cría, o valor de cría deregresado (DRP) para, μ es la media general, $\sum (SNP_{ijk})$ es la suma de todos los efectos de SNP ($k = 1,2$) a través de todos los loci (j) para el animal i . Para todos los alelos, se estima un efecto para cada uno de ellos o, de lo contrario, el efecto de sustitución de cada alelo en el locus.

Los valores de cría genómicos (GEBV) para cada animal pueden ser obtenidos como la sumatoria de los efectos de cada a :

$$GEBV_i = \sum \hat{a}$$

Existen distintos métodos con diferentes supuestos sobre la distribución *a priori* de los efectos de los SNP para ser considerados a la hora de predecir los valores de cría genómicos. Meuwissen et al. (2001) definieron dos métodos bayesianos para establecer la distribución *a priori* de la varianza de los marcadores: Bayes A utiliza una distribución Chi/cuadrado invertida para la varianza de los marcadores y Bayes B utiliza una distribución *a priori* con alta densidad de ceros para marcadores con efecto nulo.

El método GBLUP (Habier et al., 2007; Meuwissen et al., 2001; VanRaden, 2008) consiste en asumir *a priori* una distribución normal con varianza constante para los efectos de los marcadores. De acuerdo a lo propuesto por estos autores, los

valores de cría genómicos pueden estimarse directamente con un modelo simple que incluye una matriz de relaciones genómicas derivada de los genotipos y las varianzas de los efectos de los marcadores SNP (Habier et al., 2007; Nejati-Javaremi et al., 1997; VanRaden, 2008). Nejati-Javaremi et al. (1997) propusieron el uso de una matriz de relaciones alélicas como alternativa a la matriz de relaciones basada en el pedigrí para evaluar la similitud genética entre individuos. Al utilizar una matriz de relaciones realizada, la selección genómica aprovechará las desviaciones de muestreo mendeliano respecto a la matriz de relaciones promedio basada en la información del pedigrí (Goddard, 2009).

Al mismo tiempo, VanRaden (2008) definió la matriz de relaciones genómicas (G), como:

$$G = \frac{ZZ'}{k}$$

donde Z es la matriz de incidencia para los SNP con dimensión $n \times p$, de acuerdo al número de animales con genotipo (n) y número de marcadores (p), y k es un parámetro escalar $k = \text{sum}(2pq)$.

Los elementos de Z son

$$z_{ij} = \begin{cases} 0 - 2p_j & \text{si es homocigota 11} \\ 1 - 2p_j & \text{si es heterocigota 12 o 21} \\ 2 - 2p_j & \text{si es homocigota 22} \end{cases}$$

para animal i y SNP j con frecuencia alélica p_j . Como fue mencionado previamente, para todos los alelos, se estima un efecto para cada alelo o, de lo contrario, el efecto de sustitución de cada alelo en el locus. El escalar k es definido usualmente como (Habier et al., 2007; VanRaden, 2008):

$$k = 2 \sum p_j(1 - p_j)$$

lo que asume *a priori* independencia de los efectos de los SNP (Gianola et al., 2009).

Las matrices G son habitualmente singulares (VanRaden, 2008). Para permitir la inversión, VanRaden (2008) propone utilizar, en cambio, la matriz G^* , definida como:

$$G^* = w G + (1 - w) A$$

donde A es la matriz de relaciones de pedigrí y w es un parámetro de ponderación que toma valores entre 0 y 1. VanRaden (2008) sugiere utilizar un valor de $w = 0,95$.

Villanueva et al. (2005) demostraron que la selección genómica mejora la precisión de las relaciones genéticas, lo que resulta en mayores ganancias en la respuesta genética, especialmente por el uso de matriz de parentesco genómico que combina información del pedigrí con los marcadores.

En una población de animales, solo una parte de estos son genotipados, mientras que en la mayoría de los animales los valores de cría son estimados a partir de los fenotipos y los pedigrís. Para aquellos animales genotipados, los valores de cría genómicos (GEBV) son predichos a partir de sus genotipos. Esto requiere de varias etapas (Hayes et al., 2009; VanRaden, 2008): 1) evaluación tradicional con un modelo animal, 2) extracción de los pseudofenotipos por deregresar las evaluaciones o desviaciones en hijas, 3) estimar los efectos genómicos para animales genotipados utilizando modelos de padres y 4) combinar la información en un índice que considera el valor de crías promedio de los padres (PA) obtenidos de una evaluación genética y los valores de cría genómicos (Saatchi et al., 2011; VanRaden et al., 2009) como:

$$GEBV_i = b_{DGV} DGV + b_{PA} PA$$

Los modelos que se implementan en múltiples etapas asumen que no hay sesgo por selección; sin embargo, en la práctica, los individuos que son genotipados no son quienes proveen los fenotipos y, a la vez, son altamente seleccionados, lo que conduce a sesgo en las estimaciones de los VC de los animales (Vitezica et al., 2011).

2.2.4. Single step genomic BLUP (SSGBLUP)

Un método alternativo fue propuesto por Misztal et al. (2009), llamado *single-step* (ssGBLUP), y es actualmente la metodología más utilizada en las evaluaciones genéticas poblacionales que incluyen genómica. Permite combinar toda la información disponible, es decir, información genealógica a partir del pedigrí, los fenotipos (con genotipo o no) y los genotipos (con fenotipo o no) en un único paso (Aguilar et al., 2010).

Las predicciones genómicas *single step* ssGBLUP están basadas en el siguiente modelo (Legarra et al., 2009; Aguilar et al., 2010):

$$y = Xb + Wu + e$$

donde **y** es el vector de los fenotipos, **b** es el vector de los efectos fijos y **u** el vector de los efectos genéticos aditivos. **X** y **W** son las matrices de incidencia para los efectos contenidos en **b** y **u**.

Las soluciones de dicho modelo son generadas a través de las ecuaciones de los modelos mixtos ssGBLUP, obteniendo simultáneamente las predicciones genómicas de los valores de cría (**u**) y las soluciones de los efectos fijos en un único paso (Aguilar et al., 2010):

$$\begin{pmatrix} X'X & X'W \\ W'X & W'W + H^{-1}\lambda \end{pmatrix} \begin{pmatrix} b \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ W'y \end{pmatrix}$$

Donde

$$\lambda = \sigma_e^2 / \sigma_u^2, \text{ y la } \text{Var}(u) = H \sigma_u^2, \text{ y } \text{Var}(e) = I \sigma_e^2$$

Los autores proponen así un modelo similar al tradicional en el que la matriz de parentesco **A** es reemplazada por la matriz **H**, que combina tanto las relaciones de parentesco basadas en las genealogías (**A**) como la diferencia entre las relaciones basadas en genómica (**G**) y las relaciones de pedigrí. La inversa de **H** (Aguilar et al., 2010, Christensen y Lund 2010) es igual a:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-2} \end{bmatrix}$$

donde G^{-1} la inversa de la matriz de parentesco genómico, y A_{22}^{-1} , la inversa de la matriz de parentesco basados en la genealogía de animales genotipados, asumiendo una partición de la matriz A en no genotipados (1) y genotipados (2). La matriz genómica es una matriz de parentesco corregida en la medida que considera el parentesco genómico basado en marcadores en vez de pedigrís (VanRaden, 2008).

Es esperable que los modelos genómicos arrojen predicciones más precisas que las predicciones provenientes de modelos de evaluaciones convencionales BLUP, lo cual ha sido demostrado en estudios teóricos por Goddard (2009), Meuwissen (2009), en simulación por Meuwissen et al. (2001), y en datos reales (Lu et al., 2015). Esto es esperable porque a diferencia de la matriz de parentesco aditivo, la matriz de parentesco genómica captura las relaciones genómicas de la partición mendeliana en los animales evaluados, de esta manera también incrementa la conectividad entre los animales de la población de referencia y la evaluada (Pszczola et al., 2012).

Algunas de las ventajas del ssGBLUP es que considera simultáneamente la información de los animales genotipados, sus parientes y su desempeño. También permite hacer uso de fenotipos de años atrás, a pesar de no contar con los genotipos, ya que esta información es incorporada con el uso de matrices que relacionan la información genómica y la proveniente de pedigrís (Aguilar et al., 2010). El uso de las matrices que combinan todos los animales permite que las mejoras de las precisiones de los animales genotipados sean transmisibles a los parientes.

Otra de las ventajas de la metodología *single step* es que es fácilmente aplicable a modelos de análisis más complejos, como multicarácter o para características umbrales, entre otras (Legarra et al., 2014).

2.3. Predicción de valores de cría genómicos y estudio de habilidad predictiva

2.3.1. Antecedentes

Los modelos utilizados en las evaluaciones genéticas son una sobre simplificación de la realidad, pero tienen buenos resultados en poblaciones muy cercanas (Legarra y Reverter, 2018). La introducción de nuevos métodos de

evaluaciones genómicas, donde se incorpora un número importante de genotipos, lleva a la necesidad de generar herramientas para cuantificar el comportamiento de las predicciones, ya que lograr predicciones insesgadas es de suma importancia en la selección, con el fin de obtener estimaciones precisas para comparaciones entre animales y tendencias genéticas acertadas (Vitezica et al., 2011).

En las evaluaciones genéticas tradicionales basadas en pedigrí, toda la información está considerada en los fenotipos y la matriz de parentesco, sin existir sesgo por selección (Sorensen y Kennedy, 1984). Con el aumento de la selección genómica y la inclusión de genotipos de individuos que no cuentan con fenotipo, se plantea cierta preocupación sobre el impacto en la precisión de las predicciones (Aguilar et al., 2010; Mantysaari et al., 2010; Chen et al., 2013). Esto se debe a que los modelos utilizados asumen que no existe selección en los individuos que son genotipados (Hayes et al., 2009), mientras que en la práctica sí ocurre y los modelos genómicos no lo consideran (Vitezica et al., 2011). Ignorar esta información puede resultar en predicciones menos precisas y sesgadas (VanRaden et al., 2009).

El estudio de la habilidad predictiva, con o sin genómica, se refiere a la capacidad de predecir futuros fenotipos, y esto es medido a través de la precisión o *accuracy*, siendo un concepto distinto a la precisión individual de la estimación del valor de cría (Legarra y Reverter, 2018). En genética animal, para evaluar la habilidad predictiva de diferentes predictores, habitualmente se han utilizado los modelos desarrollados por Henderson (1975) a través de las predicciones insesgadas BLUP (Legarra y Reverter, 2017). Las medidas comúnmente consideradas para evaluar la calidad de las predicciones son el sesgo, la dispersión y la precisión (*accuracy*), que provienen de regresión lineal.

El sesgo se refiere a la diferencia promedio entre las predicciones (EBV, $\hat{u}$) y los verdaderos valores de cría (TBV, u). Lo deseable es que esta diferencia sea cercana a 0, lo que indicaría predicciones insesgadas.

$$\text{Sesgo: } b_0 = E(u - \hat{u})$$

La dispersión puede ser comprendida como el sesgo de la predicción y se obtiene como la pendiente de regresión (b_1) del EBV en el TBV. Tiene una esperanza

de valor de 1. Si la dispersión es mayor que 1, significa que la predicción para los animales es subestimada, mientras que, si el b1 es menor a uno, las predicciones serán sobreestimadas.

$$\text{Pendiente: } b_1 = \frac{\text{cov}(u-\hat{u})}{\text{var}(\hat{u})}$$

La precisión o *accuracy* es definida como la correlación entre el valor verdadero (TBV) y los valores de cría estimados (EBV). Una precisión cercana a 1 indica una fuerte asociación entre el TBV y los EBV estimados, mientras que valores cercanos a 0 indican bajas asociaciones. Es calculada de la siguiente manera:

$$\text{Precisión: } r = \frac{\text{cov}(u, \hat{u})}{\sqrt{\text{var}(u)\text{var}(\hat{u})}}$$

donde $\text{cov}(u, \hat{u})$ es la covarianza entre el verdadero valor de cría (u) y los valores de cría estimados ($\hat{u}$); $\text{var}(u)$, la varianza genética aditiva de la población y, $\text{var}(\hat{u})$, la varianza de los valores de cría estimados.

La precisión afecta directamente la ganancia genética, ya que $\Delta G = i r \sigma_u$ (Falconer y Mackay, 1996); por lo tanto, afecta el potencial de progreso genético. Por otro lado, podemos ver la precisión como una medida del riesgo de que el EBV del candidato cambie en la siguiente evaluación. Esta precisión es la que se obtiene de las ecuaciones de modelos mixtos y se refiere a la precisión individual (Macedo, 2021).

$$r = \sqrt{1 - \frac{PEV}{(1 + Fi)\sigma_\mu^2}}$$

donde PEV es la varianza del error de predicción; Fi , la consanguinidad del individuo i , y σ_u^2 , la varianza genética (Henderson, 1975). Los PEV (sigla en inglés) se obtienen de los elementos de la diagonal de la inversa del sistema de ecuaciones de los modelos mixtos. Esta última precisión se refiere a la precisión individual con el cual el EBV fue estimado, mientras que la anterior hace referencia a la habilidad predictiva en términos poblacionales.

El incremento del uso de información genómica en las evaluaciones genéticas, especialmente en ganado de leche, ha llevado a la selección de machos con valores de cría precisos a edades tempranas. Sin embargo, la experiencia ha demostrado algunas diferencias entre las predicciones genómicas y los resultados de las progenies *a posteriori* (Spelman et al., 2010; Sargolzaei et al., 2011). Esa diferencia, conocida como sesgo, desvío, sobreestimación o subestimación de las predicciones genómicas, puede llevar a respuestas de selección no esperadas.

En lo que refiere a predicciones genómicas en ganado de carne, las técnicas de validación también se comenzaron a implementar. El método más comúnmente utilizado para evaluar las predicciones (genómicas) son las validaciones cruzadas (VanRaden et al., 2009; Mantysaari et al., 2010; Saatchi et al., 2011). Sin embargo, otras interpretaciones desarrolladas por otros autores vienen siendo utilizadas y serán desarrolladas en este trabajo.

2.3.2. Validaciones cruzadas

La habilidad predictiva y precisiones de las predicciones genómicas son evaluadas generalmente a través de validación cruzada (VanRaden et al., 2009; Legarra et al., 2008). Esta consiste en dividir los datos en dos segmentos, un subset de datos de referencia o entrenamiento y otro set de datos de validación del modelo. El set de entrenamiento es utilizado para ajustar el modelo y el set de validación es utilizado para testear la predicción del modelo. En un mecanismo de validación alternados en varias rondas, se repite tantas veces de manera de que cada set de datos tenga la chance de ser validado (Saatchi et al., 2011). En cada ronda se omite la información de la población a validar (fenotipos) y se predicen los fenotipos a través de un modelo de predicción obtenido únicamente con la información contenida en el subset de entrenamiento.

Diferentes estrategias para elegir la población de referencia y validación o población de candidatos a selección tienen impactos directos sobre los resultados de los estudios de validación medidos a través las precisiones de las predicciones (Chen et al., 2013; Silva et al., 2016; Lu et al., 2016). Estas estrategias de división de los datos buscan analizar las consecuencias e impactos de diferentes enfoques para

conformar las poblaciones de predicción y validación y, de esta manera, evaluar el alcance y la precisión de las predicciones. Las poblaciones pueden ser divididas en función del grado de parentesco o rangos de nacimientos entre los subconjuntos de datos, al azar (Lu et al., 2016, Chen et al., 2013; Silva et al., 2016), por k-means donde separa los grupos de acuerdo a mayor distancia entre grupos y mayor similitud dentro del grupo (Saatchi et al., 2011) o por la metodología *leave one out* (Gianola y Schon, 2016), en donde en cada iteración se predice el valor del dato que se dejó fuera de la predicción. El test de validación mayormente utilizado en bases de datos pequeñas es el K-fold, el cual consiste en dividir al azar el set de datos y correlacionar las predicciones del k-set cuando los fenotipos son excluidos (Saatchi et al., 2011). Generalmente se observa que, cuanto mayor es el grado de parentesco entre la población de entrenamiento y la población de validación, se obtienen mayores precisiones (Saatchi et al., 2010; Chen et al., 2013), aunque en algunos casos puede generar sobreestimación de la habilidad predictiva. Como ejemplo, Saatchi et al. (2011) obtuvieron precisiones mayores para todas las características cuando la separación de grupos fue al azar versus cuando fue a través de k-means.

Un indicador comúnmente utilizado para medir la precisión en las validaciones cruzadas es la predictibilidad (*predictability*), la cual es definida como la correlación entre los GEBV estimados en k-fold réplicas y los valores fenotípicos ajustados, divididos por la raíz cuadrada de la heredabilidad. La predictibilidad puede ser utilizado como un estimador de la precisión (Legarra et al., 2008).

$$\text{Predictabilidad} = \text{correlación (GEBV, fenotipo corregido)} / \sqrt{h^2}$$

Los valores de cría estimados pueden haber sido estimados por BLUP, GBLUP, ssGBLUP o Bayes C, entre otros, y los valores fenotípicos ajustados o valores de cría desregresados considerados como el valor verdadero. La heredabilidad de la característica en cuestión (h^2) puede haber sido obtenida por los propios datos.

El método de validación cruzada ha sido ampliamente utilizado (Legarra et al., 2008; VanRaden et al., 2009; Mantysaari et al., 2010; Christensen et al., 2012). Sin embargo, de acuerdo a Legarra y Reverter (2018), su validez es pobremente entendida en el contexto de datos fuertemente relacionados y seleccionados.

2.3.2. Método LR (Legarra y Reverter, 2018)

Recientemente, Legarra y Reverter (2018) profundizaron sobre el estudio de la habilidad predictiva de las predicciones genómicas y analizaron las distintas estrategias para evaluarla, al mismo tiempo que observaron ciertas limitaciones de los métodos de validaciones cruzadas mencionados anteriormente.

La estrategia de validación k-fold es comúnmente aplicada para evaluar la capacidad predictiva y, como ya fue mencionado, consiste en extraer al azar distintos subsets de datos y correlacionarlos con las predicciones obtenidas de los k-fold-1 soluciones cuando los fenotipos de diferentes animales son excluidos (Saatchi et al., 2011). Sin embargo, de acuerdo a Legarra y Reverter (2017), esta metodología no es realista en el contexto de mejora animal porque no tiene en cuenta la estructura de la población, ya que es posible encontrar animales más viejos en el k-fold por predecir y animales jóvenes en el k-fold-1 fold de entrenamiento. En estas situaciones los ancestros son predichos desde las progenies, lo que no tienen sentido desde el contexto de mejora genética animal (Legarra y Reverter, 2017).

Estos autores proponen una alternativa con procedimientos semiparamétricos basada en la teoría clásica de modelos mixtos. El método LR, por “linear regression”, considera los cambios en los GEBV de una evaluación genética a la siguiente en la medida que se agrega más información y se definen ciertos parámetros que describen los cambios en la predicción de una evaluación genética vieja (*old*) y una evaluación genética reciente (*new*). También definen los dos sets de datos como el *whole data set* (*w*) o completo, que contiene todos los fenotipos, y el set parcial (*p*), que contiene los fenotipos hasta determinada fecha. En ese sentido, definen a los individuos focales como el grupo de animales jóvenes de interés por predecir que deben ser seleccionados y que disponen de información parcial hasta cierto momento. Este grupo parcial puede ser interpretado como al momento de la evaluación, cuando no se dispone de toda la información disponible. El set completo o *whole data set* (*w*) es la confirmación *a posteriori* de la correcta predicción o no de esas predicciones. Este método compara sucesivos EBV de un grupo de individuos (individuos focales), lo que permite incrementar la información con cada repetición.

Los estadísticos propuestos por Legarra y Reverter y expandidos por Bermann et al. (2021) para evaluar la calidad de las predicciones genéticas haciendo uso de las predicciones parciales (p) o totales (w) son los siguientes:

1. **Bias ($b0$):** $\hat{u}_p - \hat{u}_w$. Se refiere a la diferencia entre el valor predicho promedio y el valor verdadero (TBV, sigla en inglés), con una esperanza de que esa diferencia sea 0 para una predicción insesgada.

2. **Dispersion ($b1$):** $b_{w,p} = \frac{cov(\hat{u}_w, \hat{u}_p)}{var(\hat{u}_p)}$. Representa la dispersión de la predicción y es estimado por la pendiente de la regresión entre las predicciones entre el verdadero valor de cría estimado con el set de datos totales ($\hat{u}_w$) y el set de EBV estimados con el set de datos parcial ($\hat{u}_p$), con una esperanza de valor de 1 si la predicción no es ni sobre- ni subestimada.

3. **Reliability ($\rho^2 cov(w,p)$):** $\frac{cov(\hat{u}_w, \hat{u}_p)}{(1 - F - 2 - f) \sigma_{u,\infty}^2}$. Donde F es el coeficiente de consanguinidad; $2-f$, el parentesco promedio entre los individuos, y $\sigma_{u,\infty}^2$, la varianza genética de los individuos del set de validación.

4. **Ratio of accuracy ($\rho_{w,p}$):** $cor(\hat{u}_w, \hat{u}_p)$ definido como el coeficiente de correlación de Pearson entre el TBV y el valor de cría estimados (EBV), representados como $\hat{u}_w$ y $\hat{u}_p$, obtenido con la información parcial o completa, lo que es igual a

$$\rho_{w,p} = \frac{cov(\hat{u}_w, \hat{u}_p)}{\sqrt{var(\hat{u}_w)var(\hat{u}_p)}}$$

el valor esperado de $\rho_{w,p}$ es $\frac{acc_p}{acc_w}$, la relación de precisión obtenida entre ambos sets de datos (acc_p/acc_w).

5. Incremento relativo en precisiones por el agregado de información fenotípica al set parcial.

$$\text{Incr Phen (IP)} = \frac{1}{\rho_{w,p}} - 1$$

Valores cercanos a 1 indican que el agregado de información no mejora la precisión, mientras que valores cercanos a 0 indican un significativo incremento en las precisiones debido a la nueva información.

El método LR ha sido utilizado y testeado con datos simulados (Legarra y Reverter, 2018), en datos reales en ovinos de leche (Macedo et al., 2020) y en modelos categóricos en aves (Bermann et al., 2021).

2.4. Aspectos que afectan las precisiones de las predicciones genómicas

Las precisiones de las predicciones genómicas son determinantes del éxito de la aplicación de selección genómica, la que está determinada por múltiples aspectos (Chen et al., 2013). A través de varios estudios sobre validaciones en datos simulados y reales, se ha podido evaluar cómo ciertos aspectos afectan la habilidad predictiva de las predicciones genómicas, así como las precisiones individuales de los valores de cría estimados. Esto debería tenerse en cuenta para el diseño de las poblaciones de referencia, así como en los alcances de las predicciones genómicas bajo distintas circunstancias.

2.4.1. Tamaño y grado de conexión entre la población de entrenamiento y candidatos a la selección

El tamaño de la población de referencia es entendido como el número de animales con genotipo y fenotipo para el cual la predicción estimada es determinante para los resultados (Goddard, 2012). Van Raden et al. (2009) demostraron cómo las precisiones incrementan con el aumento del tamaño de la población de referencia, en la medida que, cuanto mayor el número de animales genotipados que dispongan del fenotipo, mejor es la estimación del efecto de cada SNP.

Está demostrado que la predicción genómica de los valores de cría para una raza es mayor cuando la población de entrenamiento es de la propia raza, en contraste con ser de otra raza o un *pool* de varias razas (Chen et al., 2013). Esto está basado en gran parte en la relación de parentesco genético y la presencia de desequilibrio de ligamiento (DL) entre la población de entrenamiento y la población de validación, en la medida que un mayor grado de parentesco o mayor DL entre la

población de entrenamiento y los candidatos va a generar una mayor precisión en los GEBV. Estudios realizados por Lu et al. (2016) en varias razas demostraron similares resultados. Saatchi et al. (2011) reporta que en Angus las predicciones genómicas son más precisas cuando los animales en la población de validación están más emparentados con la población de entrenamiento. En ganado lechero, el cual está frecuentemente muy emparentado y el tamaño efectivo de la población es bajo, el DL es mayor y las precisiones alcanzan valores más altos (Goddard, 2009).

Estudios por simulación en ganado de leche desarrollados por Pszcola et al. (2012a) también demostraron que, cuanto mayor es el grado de relación entre la población de referencia y los candidatos a la selección, mayor es la habilidad predictiva y las precisiones de predicción. Por el contrario, una fuerte relación entre la población de entrenamiento, pero una baja relación con la población por predecir tiene un efecto negativo sobre las precisiones de la predicción. Esto demostró que se debe mantener un balance entre la relación dentro de la población de entrenamiento, así como entre esta y la población por predecir. Esto es especialmente importante en poblaciones de referencia pequeñas.

Similarmente, otros autores demostraron que, en la medida que van pasando las generaciones y la población por predecir se aleja de la población de referencia, las precisiones disminuyen, lo que se asocia a una pérdida de DL (Habier et al., 2010). Para prevenir esto, es necesario mantener actualizadas las poblaciones de entrenamiento con animales provenientes de generaciones recientes, por ejemplo, animales del set de evaluación cuando los fenotipos están disponibles.

Las precisiones de las predicciones genómicas son, en general, más altas en ganado lechero que en ganado de carne u ovinos (Hayes et al., 2010). Esto no es sorprendente, ya que en ganado de leche el tamaño de las poblaciones de referencia es mayor que en ganadero de carne y ovinos, el fenotipo es medido en hijas y las heredabilidades son mayores, en contraposición a ganado de carne, cuyos fenotipos son medidos en el propio individuo. El LD es mayor en razas lecheras con tamaño de población efectiva menor y las predicciones son basadas en datos de una única raza, mientras que, en ganado de carne y ovinos, las predicciones son para una mezcla de razas (Goddard, 2012).

2.4.2. Heredabilidad de la característica

La precisión de las predicciones genómicas incrementa con la heredabilidad de la característica, ya que, a mayor heredabilidad, mayor variación de la característica es explicada por factores genéticos, los que pueden ser capturados por la información genómica (VanRaden, 2008).

Las precisiones logradas en las predicciones dependen en gran medida de la heredabilidad de la característica. En el trabajo llevado a cabo por Pszcola et al. (2012a), demostraron que, independientemente del grado de relacionamiento con la población de referencia, cuanto menor es la heredabilidad de la característica a predecir, menores serán las precisiones de las estimaciones. Al mismo tiempo, se observó que, para alcanzar un nivel de precisión aceptable cuando hay heredabilidades muy bajas, es necesario incrementar el tamaño de la población de referencia y el grado de relacionamiento de los animales por predecir con esta.

2.4.3. Panel utilizado de SNP

Las precisiones de las estimaciones dependen de la densidad de marcadores utilizados (Goddard, 2012). VanRaden et al. (2009) demostraron que paneles mayores de SNP (10 k, 20 k o 40 k) tuvieron mejores predicciones para las nueve características evaluadas en leche. Paneles más grandes determinan mejores predicciones (Meuwissen y Goddard, 2010), lo cual no parece ser un problema en la medida que los costos de los genotipados van disminuyendo (Lu et al., 2016).

2.4.4. Uso de características indicadoras en modelos multicarácter como estrategia de mejora de las predicciones

La estrategia más efectiva utilizada en los programas de mejora genética en el pasado para incrementar las precisiones de las predicciones ha sido el uso de características predictoras en modelos multivariados (Calus y Veerkamp, 2011). Los principales beneficios del uso de modelos multivariados son más evidentes para características de baja heredabilidad y aquellas con fenotipos difícil de medir (Guo et al., 2014). Estas ventajas también pueden observarse en modelos que incluyan

información genómica, donde las precisiones de las predicciones genómicas pueden verse beneficiadas por el uso de características predictoras correlacionadas.

La superioridad de los modelos multicarácter genómicos sobre modelos unicarácter ha sido demostrado en trabajos en simulaciones estocásticas (Calus y Veerkamp, 2011), simulaciones determinísticas (Calus, de Hass et al., 2013) y en datos reales (Aguilar et al., 2011; Tsuruta et al., 2011). En ganado de leche, el uso de modelos genómicos multicarácter para la predicción de tasa de concepción permitió duplicar las precisiones de las predicciones para esta característica (Aguilar et al., 2011) en comparación a modelos univariados, utilizando en ambos casos metodología SSGBLUP.

Pszczola et al. (2013) investigaron el beneficio de los modelos multivariados GBLUP para predecir el mérito genético de consumo de alimento utilizando medidas de producción de leche y peso vivo como características indicadoras. Estos estudios demostraron que, cuando se agregan observaciones en características predictoras en la población de referencia, pero fundamentalmente en la población por predecir, las precisiones (medidas por validación cruzada y por PEV, varianza del error de predicción) para la característica de difícil medición mejoran sensiblemente. Otros autores estudiaron el impacto de las correlaciones genéticas y heredabilidad sobre las respuestas a la selección utilizando los modelos multicarácter y demostraron que las respuestas obtenidas dependen de las correlaciones genéticas de la característica predictora, así como de la heredabilidad de la característica a mejorar (Calus y Veerkamp, 2011). Cuando la característica indicadora tuvo una correlación genética menor a 0,5 con la característica de interés, las precisiones logradas por incluir el fenotipo fueron menores que disponer del genotipo; pero, cuando la correlación genética fue mayor de 0,5, el uso de características correlacionadas determinó mejores precisiones para los candidatos a selección.

Estudios basados en simulaciones compararon predicciones genómicas de modelos multicarácter versus modelos unicarácter considerando diferentes heredabilidades para la característica correlacionada y distinta proporción de datos faltantes (Guo et al., 2014). Este trabajo demostró que, cuando no hubo datos faltantes, tanto los modelos multicarácter como los unicarácter presentaron similares

precisiones para las predicciones, pero, cuando hubo datos faltantes para una de las características, los modelos multivariados se comportaron mejor con mayores precisiones.

3. Evaluation of feed efficiency traits in different Hereford populations and their effect on variance component estimation

Evaluación de eficiencia de conversión de alimento en diferentes poblaciones de Hereford y su efecto en la estimación de los componentes de varianza

Resumen

Contexto. El consumo residual de alimento es una característica relevante, dado el impacto positivo en reducir los costos de alimentación y las emisiones de gases de efecto invernadero. La dificultad de disponer grandes bases de datos para esta característica es una limitante a la hora de estimar parámetros genéticos para consumo de alimento (CS) y consumo residual de alimento (RFI), por lo que la unión de distintas bases de datos ha sido una alternativa para incrementar el volumen de registros y así obtener estimaciones de parámetros genéticos más precisos. **Objetivo.** El principal objetivo de este trabajo fue comparar las bases de datos de Uruguay (URY; 780 toros) y Canadá (CAN; 1597 toros), y, evaluar la factibilidad de unir las bases de datos (ALL) para estimar parámetros genéticos para consumo (CS) y RFI. **Métodos.** El consumo de alimento y las características de crecimiento fueron medidos para ambas bases, siguiendo los protocolos establecidos por el Beef Improvement Federation. Ambas bases de datos estaban conectadas por los pedigríes, aunque dichas conexiones eran débiles. Se realizaron análisis descriptivos de los registros productivos de cada set de datos (URY, CAN) y conjunto (ALL), y los coeficientes de regresión parcial de cada regresor para CS para las tres bases de datos. Se realizaron estimaciones de componentes de varianza a través de modelos univariado o multivariados, utilizando la metodología REML, para los caracteres CS y RFI, y sus predictores, ganancia diaria, peso metabólico, espesor de grasa dorsal. **Resultados.** Existe alguna diferencia en el desempeño productivo entre los dos países ($P < 0.01$), sin embargo, no hay diferencias fenotípicas en los valores promedio de RFI entre ambas bases de datos. La estimación de heredabilidad para CS fue de 0.42 (URY), 0.41 (CAN) y 0.45 para la base conjunta (ALL), mientras que para RFI las estimaciones fueron 0.34 (URY), 0.20 (CAN) y 0.25 (ALL). Los resultados

obtenidos indican que la selección por menor RFI conduciría a una reducción en el CS, sin comprometer otras características productivas ya que las correlaciones genéticas entre RFI, ganancia y peso fueron cercanas a cero ($-0.12-0.07$).

Conclusiones. Dado que las estimaciones de parámetros genéticos fueron similares entre los sets de datos (CAN, URY), la conjunción de datos (ALL), provee estimaciones de parámetros más precisas, ya que presenta menores desviaciones estándares, especialmente en los análisis multivariados.

Implicancias. Los parámetros genéticos obtenidos en este trabajo pueden ser utilizados en las evaluaciones genéticas internacionales como nacionales.

Palabras clave: evaluación genética entre países, ganado de carne, ganado de carne, consumo de alimento, correlaciones genéticas, heredabilidad, modelo multicarácter, consumo residual de alimento, estimación de componentes de varianza.

Evaluation of feed efficiency traits in different Hereford populations and their effect on variance component estimation

Maria Isabel Pravia^{A*} , Elly Ana Navajas^A , Ignacio Aguilar^A  and Olga Ravagnolo^A 

For full list of author affiliations and declarations see end of paper

*Correspondence to:
 Maria Isabel Pravia
 Instituto Nacional de Investigación
 Agropecuaria (INIA), Estación Experimental
 Las Brujas, Ruta 48 Km. 10, Canelones,
 Uruguay
 Email: mpravia@inia.org.uy

Handling Editor:
 Sue Hatcher

ABSTRACT

Context. Residual feed intake is a relevant trait for beef cattle, given the positive impact on reducing feeding costs and greenhouse gas emissions. The lack of large databases is a restriction when estimating accurate genetic parameters for dry matter intake (DMI) and residual feed intake (RFI), and combining different data sets could be an alternative to increase the amount of data and achieve better estimations. **Aim.** The main objective was to compare Uruguayan data (URY; 780 bulls) and Canadian data (CAN; 1597 bulls), and to assess the adequacy of pooling both data sets (ALL) for the estimation of genetic parameters for DMI and RFI. **Methods.** Feed intake and growth traits phenotypes in both data sets were measured following the same protocols established by the Beef Improvement Federation. Pedigree connections among data sets existed, but were weak. Performance data were analysed for each data set, and individual partial regression coefficients for each energy sink on DMI were obtained and compared. Univariate and multivariate variance components were estimated by the restricted maximum likelihood (REML) for DMI, RFI and their energy sinks traits (average daily gain, metabolic mid weight and back fat thickness). **Key results.** There were some differences in phenotypic performance among data ($P < 0.01$); however, no differences ($P > 0.1$) were observed for phenotypic values of RFI between sets. Heritability estimates for DMI were 0.42 (URY), 0.41 (CAN) and 0.45 for ALL data, whereas heritability estimates for RFI were 0.34 (URY), 0.20 (CAN) and 0.25 for ALL data. The results obtained indicate selection on reducing RFI could lead to a decrease in DMI, without compromising other performance traits, as genetic correlations between RFI, growth and liveweight were low or close to 0 (-0.12 – 0.07). **Conclusions.** As genetic parameters were similar between national data sets (URY, CAN), pooling data (ALL) provided more accurate parameter estimations, as they presented smaller standard deviations, especially in multivariate analysis. **Implications.** Parameters estimated here may be used in international or national genetic evaluation programs.

Keywords: across country evaluation, beef cattle, feed intake, genetic correlations, heritability, multiple trait model, residual feed intake, variance component estimation.

Introduction

Improving cattle feed efficiency is of great interest for animal production systems, given the positive impact on reducing feeding costs and greenhouse gas emissions. Nevertheless, its inclusion into breeding programs has been slow due to the difficulties and high cost of measuring feed intake, which limits implementing extensive recording for this trait.

The definition of the feed efficiency-related traits that should be included in a genetic program has been discussed by many authors [e.g. Nielsen *et al.* (2013) and van der Werf (2004)]. Dry matter feed intake (DMI) is highly heritable, and it is feasible that genetic progress can be achieved by selection (Koots *et al.* 1994). However, it has a high genetic correlation with liveweight and daily weight gain, so reducing DMI would lead to reduced animal performance. Research has shown a considerable phenotypic and genetic animal variation in DMI above and below what is expected or predicted based on size and growth in Hereford (Herd and Bishop 2000) and Angus cattle (Arthur *et al.* 2001a).

Received: 24 August 2021
 Accepted: 10 June 2022
 Published: 1 August 2022

Cite this
 Pravia MI *et al.* (2022)
 Animal Production Science
 doi:10.1071/AN21420

© 2022 The Author(s) (or their
 employer(s)). Published by
 CSIRO Publishing.

This variation was defined by Koch *et al.* (1963) as residual feed intake (RFI), and is the difference between the actual feed intake, usually expressed as DMI, and the expected DMI based on observed performance and associated energy sink coefficients. These coefficients are usually estimated by least squares regression of DMI on selected energy expenditures (e.g. average daily gain, metabolic liveweight, fat deposition; Savietto *et al.* 2014). Improving RFI has the advantage of reducing DMI without compromising growth performance, as RFI is, by definition, phenotypically independent from its predictor's traits (Arthur *et al.* 2001a). Consequently, phenotypic correlations between RFI and average daily gain (ADG), metabolic weight (MWT) and back fat thickness (BFAT) are zero (Schenkel *et al.* 2004).

It is necessary to know the magnitude of a trait genetic variation to be incorporated into a breeding program, as it is the indication of the genetic improvement reachable by selection. Genetic correlations with other key production traits need to be known as well. The limitation for DMI and RFI is the availability of large databases to accurately estimate (co)variance components. One strategy is to combine databases from different countries or populations to enable more precise estimations of genetic parameters (Crowley *et al.* 2010, 2014; de Haas *et al.* 2012; Berry *et al.* 2014). However, it is necessary to study each trait in each population to see if it is possible to consider them as the same trait, as well as the associations between DMI and their energy sinks traits across environments and populations where RFI was calculated (Tempelman *et al.* 2015). Estimates of genetic correlations with important traits, such as ADG and bodyweight, are vital to avoid unfavourable correlated responses due to genetic selection on feed efficiency (Crowley *et al.* 2010).

The main objective of the present study was to estimate genetic parameters for feed intake and residual feed intake for Uruguayan Hereford cattle. As data from Uruguay are limited, we studied the performance of Uruguayan (URY) and Canadian (CAN) bulls data with the purpose of evaluating the possibility of being pooled to estimate accurate genetic parameters. Performance data were analysed for each data set, and individual partial regression coefficients for particular energy sinks on DMI were obtained and compared. Pedigree links between data sets were also analysed. Variance components for DMI, RFI and their energy sinks traits (ADG, MWT and BFAT) were estimated for each data set and the joined data set, as well as genetic and phenotypic correlation between traits.

Materials and methods

Animals and phenotypes

Uruguayan data were recorded in a large national project with the main goal of implementing genomic tools to genetically

improve feed efficiency in the Hereford breed (Navajas *et al.* 2014). Individual feed intake and growth traits were measured in 780 bulls from 78 breeders, during eight postweaning feed efficiency tests performed at Kiyú Hereford Association test station between 2014 and 2018. Bulls were registered at the Hereford Herdbook, had date of birth, birthweight and pedigree data known, and participated in the national genetic evaluation. For the feed efficiency test, bulls were at least 240 days old at the beginning of the test and no older than 390 days at the end of the test. Each breeder sent three or four young bulls belonging to the same contemporary group at weaning and sired by a minimum of two different sires.

A pre-conditioning period was defined before starting the test with the objective of allowing animals to acclimatise to others of different origin, and to the facilities and diet. This period includes: 10 days on pasture, 10 days on warming-up in a pen with traditional bunk and 15 days of warming-up to the feeding facilities. The test period was of 70 days, and the animals had ad libitum access to feed and water, to avoid data bias due to restricted feeding. Animals were fed with a total mixed ration that was formulated for postweaning growth with an average of 2.5 Mcal/kg, following the Beef Improvement Federation guidelines (Beef Improvement Federation 2010). Individual feed intake data were measured using Growsafe automated feeding system (Model 6000; GrowSafe Systems Ltd., Airdrie, AB, Canada). Average DMI for each animal was obtained as the mean daily intake on dry matter basis for a minimum of 50 valid days on test. ADG for each animal was calculated as the slope for the linear regression for all bodyweight recorded fortnightly during the test period. Only regressions estimated with coefficients of determination (R^2) > 0.95 were included. Metabolic liveweight (MWT) was the bodyweight at mid test (0.75). Rib eye area and back fat thickness (BFAT) were measured by ultrasound at the beginning, middle and end of the test, as described by Navajas *et al.* (2014) and Pravia *et al.* (2015), although only BFAT at the end of the test was used in the estimation of RFI (Ravagnolo *et al.* 2018). Applied protocols were approved by the INIA Animal Ethics Committee (INIA_2018.11).

Canadian data were provided by an agreement with the Canadian Hereford Association. This data included 1597 bulls with phenotypes, obtained in 46 tests performed from 2012 to 2016. Traits and test measurement protocols followed the BIF recommendations as in URY, which facilitated their utilisation (Crowley *et al.* 2014; Ravagnolo *et al.* 2018).

Subsets of data

Performance data of URY ($n = 780$) and CAN ($n = 1597$) were the independent data sets considered in this study. An additional data set combining information from both countries (ALL; $n = 2377$), was also included.

Phenotypic regression models for DMI

Phenotypic RFI (RFIp) was calculated for each dataset independently as the residual of the following regression model:

$$\text{DMI} = \text{CG} + b_1\text{ADG} + b_2\text{MWT} + b_3\text{BFAT} + e \quad (1)$$

where DMI is the average dry matter intake of the valid days (kg/day), CG is the contemporary group year-test-pen, ADG (kg/day) is the average daily gain obtained by linear regression of weights recorded fortnightly during the test, MWT is the metabolic weight at mid test and BFAT is ultrasound back fat thickness measure at the end of test. Coefficients b_1 , b_2 and b_3 are the partial regression coefficients of each trait on DMI, and the residual of the regression (e) represents the RFIp.

This regression model was performed using a linear model of R (RStudio Team 2020), so three RFIp (URY_RFIp, CAN_RFIp, ALL_RFIp) were obtained. Phenotypic means of the main traits recorded during the test were compared between the URY and CAN data sets. An ANOVA and Tukey test were performed to reject or accept the hypothesis that average means from different data sets are equal.

Pedigree data

Animals in both data sets had complete pedigree information up to three generations, provided by the Uruguayan Hereford Association and by the Canadian Hereford Association. Animals with records in URY and CAN data sets were sired by 322 and 369 different sires, respectively. There were three sires and 17 grandsires in common between the URY and CAN data sets.

Variance component estimation

Variance components were estimated by restricted maximum likelihood for each data subset using airmf90 (Misztal *et al.* 2002). Asymptotic standard errors of variance components and its combinations were obtained as in Meyer and Houle (2013).

Univariate models

Univariate models were defined and run for each of the three data subsets, for RFIp and traits included in its estimation (DMI, ADG, MWT and BFAT). The genetic model for the univariate analysis was:

$$y = X\beta + Zu + e \quad (2)$$

where y is the trait considered (DMI, MWT, ADG, BFAT or RFIp, resulting from Model 1), X is the incidence matrices for fixed effects (β), Z is the matrix linking animals to phenotypes, u is the vector of random animal effects and e is the vector of random residual effects. Fixed effects were

age of dam as a class effect with five levels (younger than 697 days of age; from 697 to 913 days old; from 914 to 2007 days old; older than 2008 days of age; and unknown age of dam), age of animal at test as a linear covariable and contemporary group defined by year-test-pen (1–15) for all traits, except for RFIp, as CG was considered when RFIp was calculated in Eqn 1. It was assumed that $u \sim N(0, A\sigma_u^2)$, where A is the additive relationship matrix and σ_u^2 is the genetic additive variance, and $e \sim N(0, I\sigma_e^2)$, where I is the corresponding identity matrix and σ_e^2 the residual variance.

RFIp values below -2.5 and over 2.5 were considered outliers and removed from variance component estimation.

Parameters estimated in the univariate analysis were used as priors in the bivariate and multivariate analysis for each data subset.

Bivariate and multivariate models

Bivariate analyses were performed for RFIp and DMI, and each of the other traits considered for the calculation of RFI (ADG, MWT and BFAT), to estimate the genetic and residual (co)variances. URY and CAN data were considered as the reference estimations, and compared with those estimated in the ALL data set to evaluate the similarity of genetic and phenotypic parameters, and potential improved accuracies by using a larger data set.

Six pairwise bivariate analysis were performed (RFIp with ADG, MWT and BFAT, and DMI with ADG, MWT and BFAT) for each data set. A multivariate analysis was performed also to corroborate genetic correlations of DMI and RFIp with the energy sinks (ADG, MWT, BFAT) with those obtained from bivariate analysis, and to study the genetic correlations between them.

The genetic model used for the multivariate analysis was:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} X_1\beta_1 \\ X_2\beta_2 \\ X_3\beta_3 \\ X_4\beta_4 \end{bmatrix} + \begin{bmatrix} Z_1u_1 \\ Z_2u_2 \\ Z_3u_3 \\ Z_4u_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (3)$$

where y_1 is DMI; y_2, y_3 and y_4 are ADG, MWT and BFAT; X_{1-4} and Z_{1-4} are the incidence matrices that relate data with fixed effects (β_{1-4}) and random additive genetic effects (u_{1-4}), respectively; and e is the vector of random residual effects. Fixed and animal effects and their assumption are the same as described previously in Model 2.

Results and discussion

Data description

Descriptive statistics for RFIp and all traits used in the estimation in the two independent data sets are presented

Table 1. Descriptive statistics of traits recorded in the feed efficiency tests of Uruguayan and Canadian Hereford bulls, and comparison between data sets.

	URY (n = 780)				CAN (n = 1597)			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Age (days)	283	21.6	242	333	279	29.2	171	365
ADG (kg/day)	1.43a	0.27	0.66	2.26	1.46b	0.26	0.51	2.40
MWT (kg)	75.92a	7.88	49.50	102.50	92.60b	8.10	64.70	118.61
BFAT (mm)	3.71a	1.07	1.80	8.30	4.61b	1.80	0.70	13.20
DMI (kg/day)	8.54	1.26	4.21	12.60	8.62	0.97	4.10	11.92
RFP (kg/day)	0.00	0.57	-2.68	1.89	0.00	0.56	-3.23	2.49

Different letters (a, b) in the same line indicate statistical differences between data sets ($P < 0.01$).

References: Age at beginning of test (days); ADG, average daily gain; MWT, metabolic mid weight; BFAT, back fat thickness; DMI, dry matter intake.

in Table 1. Statistical differences between means are also indicated.

As is shown in Table 1, URY bulls were lighter and leaner than CAN bulls ($P < 0.01$). There was no difference in DMI, and as expected, the RFP mean was zero in both data sets.

Phenotypic regression models for DMI

Partial regression coefficients of DMI on MWT, ADG, BFAT, intercept and adjusted coefficient of determination ($\text{adj-}R^2$) values from Model 1 are presented in Table 2 for each data set and the combined set.

All phenotypic regression models were significant ($P < 0.001$), and the higher $\text{adj-}R^2$ was 0.80 for URY data, and the lowest 0.65 for the CAN set. Regression coefficients for all traits (MWT, ADG, BFAT) were significantly different from zero ($P < 0.05$) in all data sets. Standard error deviations for partial regressions were higher for URY than for CAN. Partial regression coefficients for MWT on DMI ranged between 0.06 and 0.08 kg/day for the data sets.

Table 2. Partial regression coefficients and standard errors ($\pm$ s.e.) for energy sinks on predicted DMI for Uruguayan, Canadian and joined data set.

	URY	CAN	ALL
Intercept	0.14 (± 0.23) n.s.	0.33 (± 0.22) n.s.	-0.37 (± 0.18)***
MWT (kg)	0.08 (± 0.004)***	0.06 (± 0.002)***	0.07 (± 0.002)***
ADG (kg/day)	1.84 (± 0.12)***	1.42 (± 0.07)***	1.59 (± 0.06)***
BFAT (mm)	0.06 (± 0.02)**	0.03 (± 0.01)**	0.03 (± 0.01)**
Adj R^2	0.80	0.65	0.72
CG	15***	46***	61***
Records	780	1597	2377

Significance codes: * $P < 0.1$, ** $P < 0.01$, *** $P < 0.001$, n.s. means regression coefficient does not differ from zero ($P > 0.1$).

References: Intercept is the intercept of the regression model; Adj R^2 , the adjusted regression model; Records, the number of records considered in each data set; MWT, metabolic mid weight; ADG, average daily gain; BFAT, back fat.

ADG and BFAT regressor coefficients present some variation between both sets, pool data present middle regression for ADG, meanwhile BFAT tends to be similar for CAN. This was expected, as phenotypic values for BFAT for CAN were quite different from URY data sets. Adjusted R^2 values were similar to the 0.75 reported by Savietto *et al.* (2014), who used the same traits in the regression model. Partial regression coefficients on DMI were also similar.

RFP obtained from Model 1 (Eqn 1) were named URY_RFP, CAN_RFP and ALL_RFP according to the data set considered in the estimation. Spearman correlation coefficients among RFP from each data set with ALLdata were 0.98 and 0.99 for URY_RFP and CAN_RFP with ALL_RFP, respectively (Fig. 1a, b). Ranking of animals was not affected when different data sets were used to calculate RFP, suggesting only small changes in RFP when combining data sets, even when there were phenotypic differences between populations, as those reported here. Similar results were obtained by Crowley *et al.* (2010), who also found differences in performance traits among breeds, but RFP was estimated for the pooled data, with breed included as a fixed effect in the regression model. A similar approach was followed by Schenkel *et al.* (2004), who combined data from different breeds and test stations.

Additive genetic variances and heritability estimates

Additive genetic (VG), residual variances (VR) and heritability (h^2) estimated by univariate analysis for DMI, ADG, MWT, BFAT and RFP in URY, CAN and ALL data sets are presented in Table 3.

DMI

Genetic parameters obtained for URY and CAN were quite similar, aside from a few differences found in some of the traits. Genetic and residual variances for DMI for URY (0.42 ± 0.13) were almost twice the magnitude than for CAN (0.27 ± 0.05). When data were joined, the VG

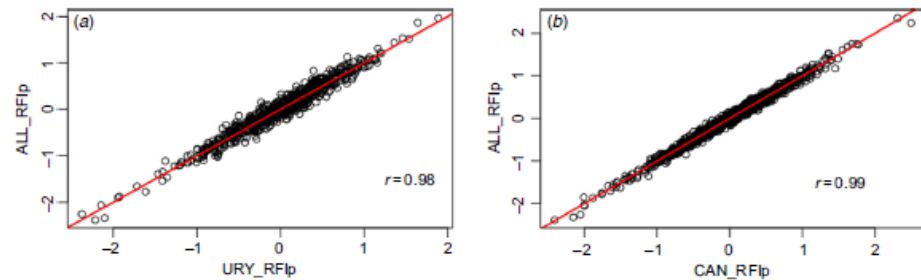


Fig. 1. (a) Correlation between RFIp obtained from URY set versus ALL data. (b) Correlation between RFIp obtained from CAN set versus ALL data set.

Table 3. Additive and residual variances, heritability, and standard errors from univariate analyses for DMI, ADG, MWT, BFAT and RFIp for Uruguayan, Canadian and joined data sets.

Trait	Parameter	URY	CAN	ALL
DMI	σ_g^2	0.42 (0.13)	0.27 (0.05)	0.35 (0.06)
	σ_e^2	0.58 (0.12)	0.39 (0.05)	0.41 (0.05)
	h^2	0.42 (0.12)	0.41 (0.07)	0.45 (0.07)
ADG	σ_g^2	0.017 (0.01)	0.020 (0.01)	0.02 (0.01)
	σ_e^2	0.026 (0.01)	0.025 (0.01)	0.030 (0.01)
	h^2	0.39 (0.12)	0.44 (0.07)	0.42 (0.06)
MWT	σ_g^2	29.2 (7.4)	28.8 (3.17)	32.7 (3.29)
	σ_e^2	18.55 (6.4)	5.58 (2.32)	6.90 (2.48)
	h^2	0.61 (0.14)	0.84 (0.07)	0.82 (0.07)
BFAT	σ_g^2	0.20 (0.10)	0.88 (0.14)	0.64 (0.09)
	σ_e^2	0.61 (0.10)	0.94 (0.12)	0.85 (0.08)
	h^2	0.24 (0.13)	0.48 (0.07)	0.42 (0.06)
RFIp	σ_g^2	0.11 (0.04)	0.06 (0.02)	0.08 (0.02)
	σ_e^2	0.21 (0.04)	0.25 (0.02)	0.23 (0.02)
	h^2	0.34 (0.13)	0.20 (0.06)	0.25 (0.06)

References: σ_g^2 , additive genetic variance; σ_e^2 , phenotypic variance; h^2 , heritability; DMI, dry matter intake; MWT, metabolic mid-weight; ADG, average daily gain; BFAT, back fat; RFIp, phenotypic residual feed intake.

presented an intermediate value (0.35 ± 0.06). Previous studies in purebred animals reported VGs for DMI ranging from 0.27 in Angus (Arthur *et al.* 2001a) to 0.60 in Charolais (Arthur *et al.* 2001b). Larger estimates were found in multibreed populations by Crowley *et al.* (2010) (0.79) and Schenkel *et al.* (2004) (0.56).

Heritability estimates obtained for DMI were equal for URY and CAN, and an equivalent estimate was obtained for the pooled data. These estimates were similar to those reported by some researchers, which ranged between 0.31 and 0.49 (Herd and Bishop 2000; Crowley *et al.* 2010).

ADG, MWT and BFAT

Variance component estimations for ADG in URY and CAN presented similar VG and VR, which leads to similar heritability (0.39 and 0.44 for URY and CAN, respectively). Estimates for MWT show similar VG between URY and CAN, although URY presents higher VR than CAN, leading to a lower h^2 (0.61 vs 0.84). BFAT parameter estimates were very different between populations. CAN presented five times more VG than URY, and twice the VR, so h^2 estimates were significantly higher than in the URY set.

Genetic variances obtained for ADG for URY and CAN populations were similar (0.017 and 0.020, respectively), and comparable with estimations of 0.014 reported by Schenkel *et al.* (2004). Heritability estimates for ADG (URY 0.39 and CAN 0.46) were slightly above those reported by Arthur *et al.* (2001a) (0.28) and Schenkel *et al.* (2004) (0.35). Genetic variances for MWT were similar between URY and CAN (29.2 and 28.8), but higher than reported by other authors in Angus cattle (Arthur *et al.* 2001a) or in a multibreed analysis (15.96, Schenkel *et al.* 2004). Heritability obtained for MWT was higher (0.61–0.87) than reported by those authors. Genetic variances and h^2 for BFAT were very different between data sets (VG: 0.20 vs 0.88, and h^2 of 0.24 vs 0.48 for URY and CAN, respectively), as was expected according to their phenotypic values. Previous studies referred to genetic variances 0.47 and h^2 of 0.35 for Angus cattle (Arthur *et al.* 2001a), an average of genetic variance of 0.70, and h^2 of 0.36 for BFAT in a multibreed analysis (Schenkel *et al.* 2004).

RFIp

Genetic variance for RFIp was higher in the URY population (0.11 ± 0.04) than in the CAN population (0.06 ± 0.02). Previous researchers obtained genetic variances that ranged between 0.11 in a multibreed analysis (Crowley *et al.* 2010), 0.14 for Angus cattle (Arthur *et al.* 2001a), 0.25 in Charolais (Arthur *et al.* 2001b) and Wagyu (Hoque *et al.* 2006) to 0.36 in a multibreed (Schenkel *et al.* 2004). RFIp had a medium heritability of $0.34 (\pm 0.13)$ in

E

URY, 0.20 (± 0.05) in CAN and 0.24 (± 0.05) in pooled data. These estimations agree with previous studies that reported h^2 for RFI of 0.16 in Hereford (Herd and Bishop 2000), 0.24 in Wagyu, 0.39 in Angus (Arthur et al. 2001a), 0.43 in Charolais (Arthur et al. 2001b) and 0.45 in multibreed analysis (Crowley et al. 2010).

The univariate analyses results indicate that URY data did not permit accurate parameter estimation, given the high standard errors observed. Similarities were found when compared with CAN estimates, except for BFAT, which encouraged combination of all data sets. The URY population benefits from the addition of Canadian data, probably due to the volume of information and genetic links, which increase the data coming from different environmental and contemporary groups. Joint population parameter estimation tends towards Canadian parameters, nevertheless, it leads to more precise parameter estimates to be used in genetic evaluations. All the traits studied were moderately heritable, and are thus possible to consider for genetic improvement programs.

Genetic and phenotypic correlations

Bivariate analyses were performed to obtain genetic and phenotypic correlations between DMI and RFIp with their predictors ADG, MWT and BFAT for URY, CAN and ALL data sets. They are presented in Tables 4 and 5, respectively. URY estimations were considered the reference, and the impact of combining information was evaluated through the joined parameter estimates and

their standard error. CAN estimations were presented to see if correlations between traits were different to URY, and thus could generate bias on estimations from combined data analysis (ALL).

When URY and CAN data sets were combined, standard errors were smaller and more accurate parameters were estimated. Results obtained from this analysis indicate more moderate genetic correlation between DMI and their regressors, ADG, MWT and BFAT in the URY data than in CAN, except for DMI and ADG, although most of them presented higher standard error deviations. The genetic correlation between DMI and ADG was strong, and ranged from 0.61 to 0.84. Similar studies also obtained high genetic correlations (0.74) between DMI and ADG in Hereford (Herd and Bishop 2000); meanwhile, other authors obtained some lower genetic correlations, but also high, 0.54 in Angus (Arthur et al. 2001a) and 0.55 in a pool of breeds (Schenkel et al. 2004; Crowley et al. 2010). Phenotypic correlations between these traits (0.53–0.69) were higher than those reported by previous studies in other breeds, which were between 0.38 (Crowley et al. 2010) and 0.47 (Arthur et al. 2001a).

Genetic correlation between DMI and MWT was 0.45 for URY, but with high s.e.. In contrast, more accurate estimates of 0.72 and 0.77 were obtained for CAN and ALL, respectively. These higher correlations agree with estimates reported in the literature. Crowley et al. (2010) estimated similar correlations (0.78) between DMI and MWT, whereas Arthur et al. (2001a, 2001b) obtained higher estimates in

Table 4. Genetic and phenotypic correlations (standard error) between DMI and ADG, MWT, BFAT for Uruguayan (URY), Canadian (CAN) and joined data sets (ALL).

	URY		CAN		ALL	
	r_g	r_p	r_g	r_p	r_g	r_p
DMI-ADG	0.84 (0.13)	0.69 (0.02)	0.61 (0.09)	0.53 (0.02)	0.68 (0.07)	0.59 (0.01)
DMI-MWT	0.45 (0.07)	0.34 (0.09)	0.72 (0.06)	0.61 (0.02)	0.77 (0.04)	0.68 (0.01)
DMI-BFAT	0.45 (0.40)	0.40 (0.03)	0.42 (0.10)	0.36 (0.02)	0.39 (0.10)	0.36 (0.02)
DMI-RFIp	0.32 (0.30)	0.55 (0.03)	0.41 (0.15)	0.69 (0.01)	0.40 (0.11)	0.63 (0.02)

References: r_g , genetic correlation; r_p , phenotypic correlation; DMI, dry matter intake; MWT, metabolic mid weight; ADG, average daily gain; BFAT, back fat; RFIp, phenotypic residual feed intake.

Table 5. Genetic and phenotypic correlation (standard error) between RFIp and ADG, MWT and BFAT for Uruguayan (URY), Canadian (CAN) and joined data sets (ALL).

	URY		CAN		ALL	
	r_g	r_p	r_g	r_p	r_g	r_p
RFIp-ADG	-0.002 (0.32)	-0.007 (0.04)	0.09 (0.18)	-0.005 (0.03)	0.07 (0.14)	-0.003 (0.02)
RFIp-MWT	-0.31 (0.25)	-0.005 (0.04)	-0.11 (0.15)	0.001 (0.03)	-0.12 (0.12)	-0.001 (0.02)
RFIp-BFAT	0.36 (0.60)	0.02 (0.02)	-0.05 (0.17)	0.02 (0.03)	-0.08 (0.13)	0.01 (0.02)

References: r_g , genetic correlation; r_p , phenotypic correlation; RFIp, phenotypic residual feed intake; MWT, metabolic mid weight; ADG, average daily gain; BFAT, back fat.

Charolais (0.83) and lower (0.65) in Angus, and Schenkel *et al.* (2004) obtained genetic correlations of 0.36 in a pool of breeds. This high correlation, as with ADG, means that it is difficult to increase MWT without increasing DMI.

Genetic correlations obtained for DMI and BFAT for the URY set were 0.45; however, they presented a high standard error (0.45 ± 0.40) compared with CAN (0.42 ± 0.10). Combined correlations were similar for all data sets, in spite of differences in phenotypic values for BFAT presented in Table 1. Previous studies reported lower correlations, 0.24–0.27 in the study with a pool of breeds and Angus (Arthur *et al.* 2001a; Schenkel *et al.* 2004).

Genetic correlation between DMI and RFI was moderate (0.32–0.40), and similarly to other traits, the combined data set allowed for more accurate parameter estimation. This correlation means efficient animals (low RFI) consume less DMI. Other studies reported higher correlation coefficients between these traits in Hereford (0.64; Herd and Bishop 2000), Angus (0.69 and 0.79; Arthur *et al.* 2001a, 2001b) and also in a multibreed analysis (0.78; Schenkel *et al.* 2004).

RFI obtained from a regression should be phenotypically independent of its component traits, but might not be genetically independent, as stated by Kennedy *et al.* (1993). The results obtained in this work confirm that the phenotypic correlation between RFI and all component traits were zero for the different set of data, as expected. Genetic correlations between RFI and their components were not statistically different from zero, but presented higher standard errors than obtained from the bivariate analysis between DMI and those traits. Estimates with ALL data were more accurate, and RFI were genetically independent from ADG (0.07), and had low and negative correlations with MWT (–0.12) and with BFAT (–0.08).

Schenkel *et al.* (2004) also found that RFI was genetically independent from ADG (0.01); however, they found a low and negative correlation with mid metabolic weight (–0.17), suggesting that selection for low RFI (efficient animals) may lead to a low increase in body size. Arthur *et al.* (2001a) reported a genetic correlation between RFI and ADG of –0.04 in Angus. Meanwhile, Herd and Bishop (2000) obtained a correlation of 0.09 in Hereford, Arthur *et al.* (2001b) of –0.10 in Charolais and Crowley *et al.* (2010) of 0.01 in a pool of breeds. Some researchers reported a negative genetic correlation of –0.21 between RFI and MWT in British breed bulls and heifers (Archer *et al.* 1997), and –0.16 in a pool of data in Ireland (Crowley *et al.* 2010); whereas Arthur *et al.* (2001a) found a nonsignificant negative correlation (–0.06) between the RFI and MWT in Angus bulls and heifers, similar to our results. Nevertheless, Herd and Bishop (2000) reported a positive correlation of 0.22 between RFI and MWT in Hereford cattle. RFI was genetically independent from BFAT, as the genetic correlations were –0.05 and –0.08 for CAN, and for the combined data set, respectively. For URY data, the genetic correlation was no different to zero, as the

estimation presented a high standard error (0.36 ± 0.6). Arthur *et al.* (2001a) reported a low, but positive, genetic correlation between RFI and BFAT (0.17) in young Angus bulls and heifers. Other authors reported similar correlations between both traits (0.16) when RFI were calculated without considering BFAT in the regression model (Schenkel *et al.* 2004). However, when BFAT was considered in the predictive regression model, genetic correlation between both traits dropped to –0.01 (Schenkel *et al.* 2004). Herd and Bishop (2000) reported a high and negative correlation between RFI and lean carcass content (–0.43), but fat was not included in the predictive model. The inclusion of BFAT in the regression model when calculating RFI has shown that reduced the genetic correlation of RFI with BFAT (Basarab *et al.* 2003; Schenkel *et al.* 2004).

Genetic covariances and correlations for DMI and predictor traits were estimated by a multivariate analysis to confirm final parameter estimates to be used in the national genetic evaluation for RFI. The results from the multivariate analysis with the ALL data set are presented in Table 6. These results confirm that additive genetic variances (0.36) and heritability (0.47) for DMI are high enough for this trait to be considered in a genetic improvement program. However, due to the inter-relationships among the traits studied, selection against DMI could have a negative impact on ADG, MWT and BFAT as a result of a high genetic correlation between DMI with ADG (0.68), MWT (0.78) and BFAT (0.39). As we expected, there was a high correlation between ADG and MWT (0.44), and moderate with BFAT (0.23). Meanwhile, there was a high correlation between MWT with BFAT (0.51). Similar correlations were found by Schenkel *et al.* (2004) for ADG–MWT (0.56), but lower for ADG–BFAT (0.14) and for ADG–BFAT (0.15).

The focus on most breeding programs is to improve overall profit, including traits related with the profit, such as liveweight and carcass weight, but at the same time controlling those traits related with the cost, mainly feed intake (Pravia *et al.* 2014). The best approach to attempt to control all traits at the same time is the implementation of

Table 6. Genetic covariances and genetic correlation between DMI and all energy sinks (MWT, ADG, BFAT) obtained from a multivariate analysis for joined data set.

	DMI (kg)	ADG (kg/day)	MWT (kg)	BFAT (mm)
DMI	0.36 (0.05)	0.06 (0.02)	2.67 (0.37)	0.19 (0.05)
ADG	0.68 (0.07)	0.02 (0.003)	0.35 (0.08)	0.025 (0.01)
MWT	0.78 (0.04)	0.44 (0.08)	32.66 (3.27)	2.33 (0.4)
BFAT	0.39 (0.09)	0.23 (0.10)	0.51 (0.07)	0.64 (0.09)

References: Additive genetic variance in diagonal; genetic covariances above and genetic correlation below the diagonal. Heritability for individual traits for ALL data were: DMI, 0.47 (0.06); ADG, 0.42 (0.04); MWT, 0.82 (0.06); BFAT, 0.43 (0.06).

selection indexes, which combines biological and economic information into one direction of improvement (Hazel 1943). The main restriction for implementing selection indexes is the lack of genetic parameters for the specific population, especially for DMI (Ponzoni and Newman 1989; Newman et al. 1992; Pravia et al. 2014). Multivariate covariance components between DMI and growth traits obtained here from joined data could be used as the reference parameters in future analysis. This should allow the implementation of selection indices and to establish an alternative computation of RFI that considers genetic covariances between traits, as suggested by some authors (Kennedy et al. 1993; MacNeil et al. 2011). This alternative genetic RFI is obtained directly as a linear combination from each genetic breeding value for energy sinks traits in a selection index.

Implications

The lack of large databases is a restriction for accurate genetic parameters for DMI and RFI, and combining different data sets could be an alternative to increase the amount of data and achieve better estimations. Several authors combine information from different breeds (Crowley et al. 2010), countries (Manzanilla-Pech et al. 2016) and test stations (Schenkel et al. 2004; Savietto et al. 2014) to estimate genetic parameters or for genomic predictions. Most did not focus on discussing the reliability of combining different databases for feed efficiency traits, but tried to include those effects in the model.

In this work, the main objective was to compare different populations and the adequacy of pooling data for the estimation of genetic parameters. Our study showed that estimates of genetic parameters for feed efficiency traits were similar between URY and CAN data sets without relevant inconsistencies. This facilitates joining both databases, and pooling the data (URY-CAN) provided the most accurate genetic parameters for DMI and RFI, and may be used in international genetic evaluation programs.

Results from the combined data indicate that genetic variation in DMI exists, and that responses to selection should be possible, as h^2 was high (0.45). Estimated genetic parameters obtained here agreed with other estimates (Herd and Bishop 2000; Arthur et al. 2001a; Schenkel et al. 2004; Crowley et al. 2010). It needs to be pointed out that due to the interrelationships among the traits studied, selection against DMI could have a negative impact on ADG, MWT and BFAT as a result of high genetic correlations between DMI with ADG (0.77), MWT (0.68) and BFAT (0.40).

According to our results, genetic variation on DMI above and below what is expected based on MWT, ADG and BFAT was observed, and genetic improvement on RFI is possible, as it has a moderate h^2 of 0.25 and was in agreement with

parameters reported in the literature. The results obtained in our study indicate selection on reducing RFI could lead to on lower DMI, without compromising other performance traits, as genetic correlations between RFI and growth and liveweight were low or close to 0 (−0.12–0.07).

Conclusions

This work demonstrates the viability of collecting information from different sources to create a larger database that will allow estimating accurate genetic parameters, and facilitate future genetic and genomic analysis. There were small differences in phenotypic performance among databases from different origins; however, no differences were observed for RFI.

When data were combined, an improvement in parameter estimation and more accurate estimations for genetic variances and genetic correlations between DMI and RFI with their predictor traits (ADG, MWT and BFAT) were obtained.

Genetic variances for DMI and RFI were estimated, and presented medium-to-high heritability, which indicate genetic responses to selection should be possible. The parameter estimations obtained were in accordance with information previously reported by other researchers. This result suggests that combining data from different sets is a good alternative to generate better parameter estimations for scarcely recorded traits that could be used for genetic evaluations, or for selection recommendations.

References

- Archer JA, Arthur PF, Herd RM, Parnell PF, Pitchford WS (1997) Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *Journal of Animal Science* **75**, 2024–2032. doi:10.2527/1997.7582024x
- Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF (2001a) Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *Journal of Animal Science* **79**, 2805–2811. doi:10.2527/2001.79112805x
- Arthur PF, Renand G, Kraus D (2001b) Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. *Livestock Production Science* **68**, 131–139. doi:10.1016/S0301-6226(00)00243-8
- Basarab JA, Price MA, Aalhus JL, Okine EK, Snelling WM, Lyle KL (2003) Residual feed intake and body composition in young growing cattle. *Canadian Journal of Animal Science* **83**, 189–204. doi:10.4141/A02-065
- Berry DP, Coffey MP, Pryce JE, de Haas V, Lovendahl P, Krattenmacher N, Crowley JJ, Wang Z, Spurlock D, Weigel K, Macdonald K, Veerkamp RF (2014) International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *Journal of Dairy Science* **97**, 3894–3905. doi:10.3168/jds.2013-7548
- BLF (2010) Guidelines for uniform beef improvement programs. 9th edn, Beef Improvement Federation, North Carolina State University, Raleigh, 24–27.
- Crowley JJ, McGree M, Kenny DA, Crews DH Jr., Evans RD, Berry DP (2010) Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. *Journal of Animal Science* **88**, 885–894. doi:10.2527/jas.2009-1852

- Crowley JJ, Stothard P, Basarab J, Miller SP, Li C, Wang Z, Plastow G, de Pauw M, Moore SM, Lu DB (2014) Collation of data and genetic parameter estimation in different experimental Canadian beef cattle populations measured for feed efficiency. In 'Proceedings 10th world congress of genetics applied to livestock production' (WCGALP). Paper 117, Vancouver, Canada. Available at <https://asas.confex.com/asas/WCGALP14/webprogram/Paper9209.html>
- de Haas Y, Calus MPL, Veerkamp RF, Wall E, Coffey MP, Daetwyler HD, Hayes BJ, Pryce JE (2012) Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. *Journal of Dairy Science* **95**, 6103–6112. doi:10.3168/jds.2011-5280
- Hazel LN (1943) The genetic basis for constructing selection indexes. *Genetics* **28**(6), 476–490. doi:10.1093/genetics/28.6.476
- Herd RM, Bishop SC (2000) Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livestock Production Science* **63**, 111–119. doi:10.1016/S0301-6226(99)00122-0
- Hoque MA, Arthur PF, Hiramoto K, Oikawa T (2006) Genetic relationship between different measures of feed efficiency and its component traits in Japanese Black (Wagyu) bulls. *Livestock Science* **99**, 111–118. doi:10.1016/j.livprodsci.2005.06.004
- Kennedy BW, van der Werf JH, Meuwissen TH (1993) Genetic and statistical properties of residual feed intake. *Journal of Animal Science* **71**, 3239–3250. doi:10.2527/1993.71123239x
- Koch RM, Swiger LA, Chambers D, Gregory KE (1963) Efficiency of feed use in beef cattle. *Journal of Animal Science* **22**, 486–494. doi:10.2527/jas1963.222486x
- Koots KR, Gibson JP, Smith C, Wilton JW (1994) Analysis of published genetic parameter estimates for beef production traits. 1. Heritability. *Animal Breeding Abstract* **62**, 309–338.
- MacNeil MD, Lopez-Villalobos N, Northcutt SL (2011) A prototype national cattle evaluation for feed intake and efficiency of Angus cattle. *Journal of Animal Science* **89**, 3917–3923. doi:10.2527/jas.2011-4124
- Manzanilla-Pech CIV, Veerkamp RF, Tempelman RJ, van Pelt ML, Weigel KA, VandeHaar M, Lawlor TJ, Spurlock DM, Armentano LE, Staples CR, Hanigan M, De Haas Y (2016) Genetic parameters between feed-intake-related traits and conformation in 2 separate dairy populations—the Netherlands and United States. *Journal of Dairy Science* **99**, 443–457. doi:10.3168/jds.2015-9727
- Mayer K, Houle D (2013) Sampling based approximation of confidence intervals for functions of genetic covariance matrices. In 'Proceedings of the Twentieth Conference of the Association for the Advancement of Animal Breeding and Genetics, Translating Science into Action, 20th–23rd October 2013, Napier, New Zealand'. pp. 523–526. ref. 10.
- Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH (2002) BLUPF90 and related programs (BGF90). In 'Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002'. Session 28 2002. pp. 1–2. ref. 7.
- Navajas EA, Pravia MI, Lema M, Clariget J, Aguilar I, Ravagnolo O, Brito G, Peraza P, Dalla Rizza M, Montossi F (2014) Genetic improvement of feed efficiency and carcass and meat quality of Hereford cattle by genomics. In 'Proceedings of the 60th international congress meat science technology, Punta del Este, Uruguay'. pp. 17–22.
- Newman S, Morris CA, Baker RL, Nicol GB (1992) Genetic improvement of beef cattle in New Zealand: breeding objectives. *Livestock Production Science* **32**, 111–130. doi:10.1016/S0301-6226(12)80031-5
- Nielsen MK, MacNeil MD, Dekkers JCM, Crews DH Jr., Rathje TA, Enns RM, Weaber RL (2013) Review: Life-cycle, total-industry genetic improvement of feed efficiency in beef cattle: blueprint for the Beef Improvement Federation. *The Professional Animal Scientist* **29**, 559–565. doi:10.15232/S1080-7446(15)30285-0
- Ponzi RW, Newman S (1989) Developing breeding objectives for Australian beef cattle production. *Animal Science* **49**, 35–47. doi:10.1017/S0003356100004232
- Pravia MI, Ravagnolo O, Urioste JI, Garrick DJ (2014) Identification of breeding objectives using a bioeconomic model for a beef cattle production system in Uruguay. *Livestock Science* **160**, 21–28. doi:10.1016/j.livsci.2013.12.006
- Pravia MI, Ciappesoni G, Aguilar I, Ravagnolo O, Lema M, Navajas EA (2015) Analysis of feed intake and feed efficiency in Hereford Bulls and steers. In 'Proceeding 52a Reunião Anual da Sociedade Brasileira de Zootecnia, Belo Horizonte, MG, 19–23 July'.
- Ravagnolo O, Aguilar I, Crowley JJ, Pravia MI, Lema M, Macedo FL, Scott S, Navajas EA (2018) Accuracy of genomic predictions of residual feed intake in Hereford with Uruguayan and Canadian training populations. In 'Proceedings of 11th world congress on genetics applied to livestock production (WCGALP), Auckland, New Zealand'. p. 723. Available at <http://www.wcgalp.org>
- RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. Available at <http://www.rstudio.com/>
- Saviotto D, Berry DP, Priggens NC (2014) Towards an improved estimation of the biological components of residual feed intake in growing cattle. *Journal of Animal Science* **92**, 467–476. doi:10.2527/jas.2013-6894
- Schenkel FS, Miller SP, Wilton JW (2004) Genetic parameters and breed differences for feed efficiency, growth, and body composition traits of young beef bulls. *Canadian Journal of Animal Science* **84**, 177–185. doi:10.4141/A03-085
- Tempelman RJ, Spurlock DM, Coffey M, Veerkamp RF, Armentano LE, Weigel KA, de Haas Y, Staples CR, Connor EE, Lu Y, VandeHaar MJ (2015) Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *Journal of Dairy Science* **98**, 2013–2026. doi:10.3168/jds.2014.8510
- van der Werf JHJ (2004) Is it useful to define residual feed intake as a trait in animal breeding programs?. *Australian Journal of Experimental Agriculture* **44**, 405–409. doi:10.1071/EA02105

Data availability. The data that support this study will be shared upon reasonable request to the corresponding author.

Conflicts of interest. The authors declare no conflicts of interest.

Declaration of funding. Financial support was provided by the National Agency for Research and Innovation (grant RTS_I_2012_I_3489).

Acknowledgements. We thank the institution and their staff for their significant contribution to this collaborative project: Uruguayan Hereford Breeders Association, the Ministry of Livestock, Agriculture and Fishery, the National Meat Institute, the Rural Association of Uruguay and the Institute of Biological Research Clemente Estable. We are also grateful with the Canadian Hereford Association.

Author affiliation

[†]Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental Las Brujas, Ruta 48 Km. 10, Canelones, Uruguay.

4. Alternative models to predict residual feed intake in Hereford breed and effects on their breeding value accuracy

Modelos alternativos para predecir el consumo residual de alimento en la raza Hereford y sus efectos sobre las precisiones de los valores de cría

Resumen

El consumo residual de alimento (RFI) es una característica de relevancia económica, pero costosa de medir lo que lleva a que un bajo número de los candidatos a la selección son evaluados y disponen de estimaciones de mérito genético (EPD) precisos. El modelo habitualmente considerado para la evaluación genética de RFI consiste en un modelo unicarácter, donde la variable RFI es el residuo de una regresión lineal entre el consumo y las características medidas en la prueba de consumo de alimento. Este estudio evaluó dos aproximaciones alternativas para predecir el mérito genético para RFI. En primer lugar, la combinación de características medidas en la prueba de consumo en un modelo multivariado para derivar en una medida de RFI a través de un índice de selección; en segundo lugar, la inclusión del peso al destete (PD) como característica indicadora combinada con el modelo anterior. Las estimaciones de mérito genético (EPD) para las dos alternativas propuestas fueron equivalentes a las estimaciones de RFI convencional. El uso de PD como característica predictora en la población de referencia y candidatos a predecir puede ser una manera económica para incrementar las precisiones de los EPDs. Esto, además, permitiría eliminar el posible sesgo que se da por la preselección de animales que participan de pruebas de consumo, y permitía extender la predicción a mayor número de candidatos a la selección.

48. Alternative models to predict residual feed intake in Hereford breed and effects on their breeding values accuracy

M.I. Pravia*, E.A. Navajas, I. Aguilar and O. Ravagnolo

*Instituto Nacional de Investigación Agropecuaria, Las Brujas, Rincón del Colorado, Canelones, Uruguay;
mpravia@inia.org.uy*

Abstract

Residual feed intake (RFI) is a relevant trait, but expensive to measure so few candidates are phenotyped and have accurate GEBVs. Genetic evaluation corresponds to a single trait evaluation, where RFI are the residual from a lineal regression on feed intake on performance traits during feed-test. This study evaluates two approaches to predict a GEBV for RFI. First combining feed performance traits in a multitrait model to calculate a single measure of RFI using selection index, and 2nd the inclusion of weaning weight (WW) as an indicator trait combined with previous model. The GEBVs for the alternatives proposed were equivalent to conventional RFI. Using WW as a predictor trait on reference animals and candidates could be an inexpensive way to increase the accuracy of GEBVs. It allows to remove possible bias due to preselection of animals that participate on feed test and to extend prediction to more selection candidates.

Introduction

Residual feed intake (RFI) has been included in breeding programs with the aim of reducing feed intake without compromising animal performance. Because it is an expensive and labor-intensive to measure trait, only a small percentage of selection candidates are phenotyped. Traditional genetic predictions for the rest of candidates relies on pedigree relationships with animals with records and have low accuracies. Although genomic selection is a promising tool for difficult to measure traits, accuracies depend on reference populations sizes. Using easily recordable predictor traits on both reference animals and candidates without RFI phenotypes could be another inexpensive way to increase accuracy. This approach also allows to remove possible bias due to preselection of animals that participate on feed test and to extend predictions to other selection candidates.

Conventional genetic evaluation of RFI is a two-step process. First, RFI is obtained as the residual from a lineal regression for DMI that includes metabolic liveweight (MWT), average daily gain (ADG), and ultrasound back fat (UBF) measured in feed intake tests. Breeding values are then estimated based on this phenotypic RFI in a genetic univariate model (Ravagnolo *et al.*, 2018). An alternative approach to calculate genetic merit for RFI would be by combining feed intake, growth and backfat in a multi-trait model to calculate a single measure of RFI using selection index theory (Kennedy *et al.*, 1993). Another approach proposed in this work is to use weaning weight (WW) as an indicator trait, which is already available and easily recordable in all animals, combined with DMI, ADG, MWT and UBF in a multi-trait model to calculate RFI using selection index as in previous model. The objective of this work is to analyze and compare the new approaches with the conventional methodology used in the Hereford breed in Uruguay.

Materials & methods

Animals and data. Phenotypic DMI, ADG, MWT and UBF data of 780 Hereford bulls were recorded in 9 postweaning 70-day feed intake tests, using an automated system (Growsafe) for DMI recording and following Beef Improvement Federation guidelines (BIF, 2010). Three generations of pedigree (26,572 animals), as well as genomic information of 5,439 animals imputed to the 50k SNP (BovineSNP50, Illumina San Diego, CA), were included in the analysis.

R.F. Veerkamp and Y. de Haas (eds) 

Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)

DOI: 10.3920/978-90-8686-840-4_48 © M.I. Pravia *et al.* 2022

240

Methods. Three models were used for the estimation of genomic breeding values (GEBV) for RFI. In Model 1, RFI was first obtained as the residual of the following regression model:

$$\text{DMI} = \text{CG} + b_1 \text{ADG} + b_2 \text{MWT} + b_3 \text{UBF} + \text{RFI} \quad (1)$$

where CG is the feed intake test contemporary group (Test×Pen, n-15), DMI, ADG, MWT and UBF were defined previously, and b_1, b_2, b_3 the partial regression coefficient of each trait on DMI. Then, GEBV for RFI (RFI_{g1}) were obtained using GBLUP as implemented in BLUPF90 (Misztal *et al.*, 2016).

The genetic model was:

$$Y = Xb + Zu + e \quad (2)$$

where Y is RFI obtained from (1), X is the incidence matrix for fixed effects (age of the animal in days at start of test as a covariable, age of dam as five class variable), Z is the matrix linking animals to phenotypes, u is the vector of genetic breeding values, and e is the vector of residual effects. It was assumed that $u_{\text{RFI}} \sim N(0, H\sigma_u^2)$ and $e \sim N(0, I\sigma_e^2)$, where H is the combination of pedigree-based relationships and differences between pedigree-based and genomic-based relationships (Aguilar *et al.*, 2010).

Model 2 is a 4-multi-trait model which included all traits used for RFI calculation (DMI, ADG, MWT and UBF):

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} X_1\beta_1 \\ X_2\beta_2 \\ X_3\beta_3 \\ X_4\beta_4 \end{bmatrix} + \begin{bmatrix} Z_1u_1 \\ Z_2u_2 \\ Z_3u_3 \\ Z_4u_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (3)$$

where Y_{1-4} are DMI, ADG, MWT, UBF, and X and Z are the incidence matrices previously described. Solutions for GEBV for each trait were obtained, and the estimated breeding value for RFI (RFI_{g2}) was calculated using selection index theory as follows:

$$\text{RFI}_{g2} = u_1 - b_2u_2 - b_3u_3 - b_4u_4 \quad (4)$$

where 1-4 are the GEBVs solutions for DMI (u_1), ADG (u_2), MWT (u_3), UBF(u_4) predicted from the model [3], and the RFI index coefficients calculated as:

$$V^{-1}C = \begin{bmatrix} b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad V = \begin{bmatrix} \sigma_2^2 & \sigma_{u2u3} & \sigma_{u2u4} \\ \sigma_{u2u3} & \sigma_3^2 & \sigma_{u3u4} \\ \sigma_{u2u4} & \sigma_{u3u4} & \sigma_4^2 \end{bmatrix}, \quad C = \begin{bmatrix} \sigma_{u1u2} \\ \sigma_{u1u3} \\ \sigma_{u1u4} \end{bmatrix} \quad (5)$$

where V are the genetic variances and covariances for ADG (σ_2^2), MWT(σ_3^2), and UBF (σ_4^2), C the genetic covariances between DMI with ADG (σ_{u1u2}), with MWT (σ_{u1u3}), and with UBF (σ_{u1u4}) (Table 1a), which were estimated using AIREML (Misztal *et al.*, 2016)

For Model 3, WW of the 780 animals with DMI, ADG, MWT and UBF data and 4,930 animals presented in their 235-contemporary groups (only with WW) at weaning were included in a 5-multi-trait model to estimate genetic (co)variances. Representation matrix is similarly to model (5). Genetic variances and covariances estimates are presented in Table 1(b). Solutions for the five traits were obtained, but only GEBVs for DMI, ADG, MWT and UBF were combined as in (4) for the estimation of RFI_{g3}.

Table 1. Genetic co-variance components and genetic correlations obtained for 4-trait model used in RFI₂ (a), and 5-trait model used in RFI₃ (b).

a)	DMI (kg)	ADG (kg/d)	MWT (kg)	UBF (mm)	b)	WW (kg)	DMI (kg)	ADG (kg/d)	MWT (kg)	UBF (mm)
					WW	249.6	6.65	1.47	72.46	3.24
DMI	0.36	0.06	2.67	0.19	DMI	0.54	0.61	0.11	4.63	0.24
ADG	0.68	0.02	0.35	0.03	ADG	0.57	0.83	0.03	0.88	0.04
MWT	0.78	0.44	32.66	2.33	MWT	0.66	0.85	0.78	48.43	2.22
UBF	0.39	0.23	0.51	0.64	UBF	0.40	0.61	0.43	0.62	0.26

Additive genetic variance in diagonal, genetic covariances above and genetic correlations below the diagonal.

Accuracy. Individual GEBV accuracies were calculated using prediction error variances (PEV) from the inverse of the mixed model equations, as: $r = \sqrt{1 - \text{PEV}/(1+F)\sigma_g^2}$, where σ_g^2 is the genetic variance of RFI₁ ($\sigma_g^2 = 0.08$). For multi-trait analysis, the accuracies for RFI₂ and RFI₃ were calculated using PEV and PEC (prediction error covariances) (Tier and Meyer, 2004).

Bias and re-ranking. Slopes of regression and Spearman rank correlations between the alternative GEBV obtained from the different models were computed to investigate bias and changes in the ranking of animals.

Results & discussion

The high correlation coefficient and the regression slope of 1 between RFI₁ and RFI₂ (Figure 1a) indicated that Models 1 and 2 are equivalent, as described by Kennedy *et al.* (1993). The comparison of GEBVs by Model 1 and Model 3 showed some changes in ranking (Figure 1b), although the correlation between both was high. The regression slope in this case was above one indicating an underestimation of GEBVs when conventional Model 1 is used, which is corrected by Model 3. Based on Pszczola *et al.* (2013), we can assumed that RFI₃ are less biased than RFI₁ due to: (1) it enables the model to more accurately predict the level of the breeding value of an evaluated animal relative to the reference animals because records of own phenotypes for predictor traits on the evaluated animals give information about the Mendelian sampling, and (2) it accounts for differences between the evaluated animals (animals with scarcely trait records) and the reference population where all animals had the predictor trait record.

Higher standard deviations were observed for RFI₂ (0.18) and RFI₃ (0.23), in comparison with RFI₁ (0.17). This should be consequence of the genetic variances components of all traits considered in the multi-trait model. The mean of RFI₂ (0.016), and RFI₃ (0.021) were higher than RFI₁ (0.014), as result of the underestimation of predicted GEBVs mentioned above.

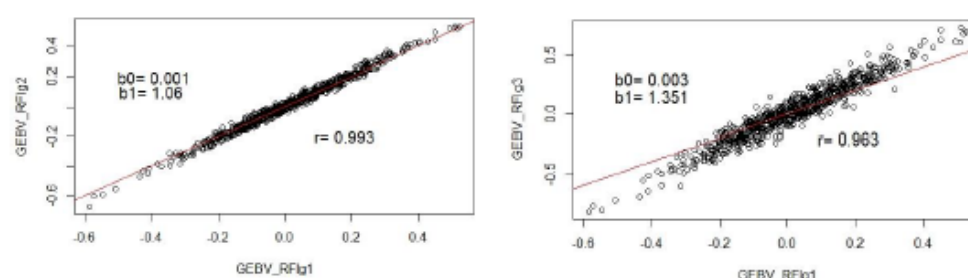


Figure 1. Correlations (r), intercept (b_0) and slope of regression (b_1) between alternatives GEBVs for RFI for the group of animals with phenotypes. (a) RFI₁ vs RFI₂, (b) RFI₁ vs RFI₃.

5. Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake

Habilidad predictiva de un modelo genómico multicarácter para la evaluación de consumo residual de alimento

Resumen

La selección por eficiencia alimenticia es el objetivo de muchos programas de mejoramiento genético en ganado de carne. El consumo residual de alimento (RFI, por sus siglas en inglés) es la característica que se ha incluido en las evaluaciones genéticas por reducir el consumo de alimento sin comprometer los rasgos de desempeño, como el peso vivo, la ganancia de peso o las características de la canal. Sin embargo, la medición de consumo de alimento es costoso por lo que un bajo porcentaje de los candidatos a la selección son medidos. La selección genómica se ha convertido en una herramienta muy importante para obtener progresos genéticos efectivos en este tipo de caracteres. Otra estrategia eficaz ha sido la implementación de la predicción a través de modelos multicarácter incorporando caracteres correlacionados de fácil registro tanto en animales de referencia como en candidatos sin fenotipos, lo que podría ser una forma económica de aumentar la precisión. El objetivo de este trabajo fue analizar y comparar la capacidad predictiva de dos enfoques alternativos para predecir el mérito genético para RFI (GEBVs) basados en información genómica. La población de estudio consistió en toros Hereford de Uruguay candidatos a la selección, que contaban con genotipado disponible. El primer modelo analizado fue el modelo univacaracter convencional para RFI, y el segundo modelo fue un modelo multicarácter que incluyó una característica predictora como el peso al destete (PD), además de los caracteres utilizados en el primer modelo (consumo de materia seca (CS), peso metabólico promedio (PM), ganancia media diaria en prueba (GAN) y grasa dorsal medida por ultrasonido (EGD)). Los GEBVs obtenidos del modelo multicarácter se combinaron utilizando la teoría de índices de selección para derivar los valores de cría para RFI. Todos los análisis se realizaron utilizando el procedimiento ssGBLUP. La habilidad predictiva

de ambos modelos se probó utilizando dos estrategias de validación: 30 réplicas diferentes de grupos aleatorios de animales, y validación a través de 9 pruebas diferentes. La capacidad predictiva de los modelos se evaluó a través de los siguientes parámetros: sesgo, dispersión, relación de precisiones y el aumento relativo en la precisión al agregar información fenotípica. Todos los parámetros mostraron que el modelo unicarácter superó al modelo multicarácter, independientemente de la estrategia de validación considerada. Estos resultados indican que la inclusión del PD como carácter predictivo en un análisis multicarácter no mejora la capacidad de predicción para el mérito genético de RFI cuando todos los animales a predecir están genotipados. **Palabras clave:** predicción genómica, modelo multicarácter, estrategias de validación.

Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake

Maria Isabel Pravia  | Elly Ana Navajas  | Ignacio Aguilar  | Olga Ravagnolo 

Instituto Nacional de Investigación
Agropecuaria, INIA Uruguay,
Canelones, Uruguay

Correspondence

Maria Isabel Pravia, Instituto Nacional
de Investigación Agropecuaria, INIA
Las Brujas, Canelones, Uruguay.
Email: mpravia@inia.org.uy

Abstract

Selection for feed efficiency is the goal for many genetic breeding programs in beef cattle. Residual feed intake has been included in genetic evaluations to reduce feed intake without compromising performance traits as liveweight, body gain or carcass traits. However, measuring feed intake is expensive, and only a small percentage of selection candidates are phenotyped. Genomic selection has become a very important tool to achieve effective genetic progress in these traits. Another effective strategy has been the implementation of multi-trait prediction using easily recordable predictor traits on both reference animals and candidates without phenotypes, and this could be another inexpensive way to increase accuracy. The objective of this work was to analyse and compare the prediction ability of two alternative different approaches to predict GEBVs for RFI. The population of inference was Hereford bulls in Uruguay that were genotyped candidates for to selection. The first model was the conventional univariate model for RFI and the second model was a multi-trait model which included a predictor trait (weaning weight, WW), in addition to the traits used in the first one (dry matter intake, metabolic mid test weight, average daily gain and ultrasound back fat) (DMI, MWT, ADG, UBF, respectively). GEBVs from the multi-trait model were combined using selection index theory to derive RFI values. All analyses were performed using ssGBLUP procedure. The prediction ability of both models was tested using two validation strategies (30 different replicates of random groups of animals and validation across 9 different feed intake tests). The prediction quality was assessed by the following parameters: bias, dispersion, ratio of accuracies and the relative increase in accuracy by adding phenotypic information. All parameters showed that the univariate model outperforms the multi-trait model, regardless of the validation strategy considered. These results indicate that including WW as a proxy trait in a multi-trait analysis does not improve the prediction ability when all animals to be predicted are genotyped.

KEYWORDS

genomic prediction, multi-trait, validation strategies

1 | INTRODUCTION

Residual feed intake (RFI) has been included in breeding programs with the aim of reducing feed intake without compromising animal performance. Because it is an expensive and labour-intensive trait to measure, only a small percentage of selection candidates are phenotyped. Traditional genetic predictions for other candidates rely on pedigree relationships with phenotyped animals and tend to have low accuracies. Genomic selection is an important tool for traits that are difficult or expensive to measure (Hayes et al., 2009). However, prediction accuracies depend on the size and design of the reference populations used (Pszczola, Strabel, Mulder, & Calus, 2012). One effective strategy used to deal with such traits in the past, without the use of genotypes, has been the implementation of multi-trait prediction with correlated traits that are easier to record (Calus & Veerkamp, 2011). Benefits have been also reported in multi-trait genomic models. Using easily recordable predictor traits in both reference animals and candidates without phenotypes could be another inexpensive way to increase accuracy (Guo et al., 2014; Manzanilla-Pech et al., 2017; Pszczola et al., 2013).

Conventional genetic evaluation of RFI is a two-step process. First, RFI is obtained as the residual from a linear regression for dry matter intake (DMI) that includes metabolic liveweight (MWT), average daily gain (ADG) and ultrasound back fat (UBF) measured in feed intake tests. Breeding values are then estimated based on this phenotypic RFI in a genetic univariate model (Basarab et al., 2003; Ravagnolo et al., 2018). An alternative approach to predict the genetic merit for RFI would be combining DMI, MWT, ADG and UBF in a multi-trait model to calculate a single measure of RFI using selection index theory (Kennedy et al., 1993). Another approach is using an indicator trait as weaning weight (WW), which is easily recordable and already available in selection candidates. This indicator trait can be combined with DMI, ADG, MWT and UBF in a multi-trait model to calculate RFI using selection index. This may lead to less biased and more accurate genomic estimated breeding values (GEBVs) than those obtained from a univariate model that only includes measures from animals with phenotypic feed intake (MacNeil et al., 2011). A previous study demonstrated that GEBVs obtained from the univariate analysis were equivalent to those obtained from the multi-trait SSGBLUP (Pravia et al., 2022b). Although higher individual GEBV accuracies were attained from the multi-trait approach due to the utilization of a highly heritable and strongly correlated trait (WW), it was still necessary to investigate the predictive ability of this approach to assess its potential benefits.

Prediction ability in genomic selection refers to the capacity to predict future phenotypes and is measured by population accuracy, which is a distinct concept from individual accuracy (Legarra & Reverter, 2018). The method most used to evaluate genomic predictions is cross-validation analysis (Mäntysaari et al., 2010; Saatchi et al., 2011; VanRaden et al., 2009). This test relies on comparing GEBV or predicted phenotypes to (precorrected) observed phenotypes, deregressed proofs or yield deviations or comparing GEBV to highly accurate EBV obtained from progeny tests. The k-fold validation strategy is commonly applied to small data sets and consists of randomly taking several data subset and correlating the predictions obtained from the k-fold when phenotypes for this fold excluded (Saatchi et al., 2011). However, this method does not take into account population structure, hence it is possible to have old animals in fold k and young animals in fold $k-1$. In that situation, ancestors are predicted from progeny, which does not make sense in the animal breeding context (Legarra & Reverter, 2017). These authors proposed an alternative approach with semiparametric procedures based on the classical theory of genetic evaluation on mixed model equations. The lineal regression (LR) approach considers the changes in EBVs from one genetic evaluation to the next one, when new data have been added (Reverter et al., 1994) and the definition of a group of parameters, which describes the change of predictions from "old" to "recent" evaluations that can be used to derive estimators for bias, dispersion and accuracy to evaluate the quality of genetic predictions. The LR method focus on the estimates of "population" accuracy, which is the correlation between the true breeding value (TBV) and the estimated breeding value (EBV) across individuals in a population (Legarra & Reverter, 2018). These authors propose some parameters to evaluate and compare the predictive ability of the models and to maximize genetic progress.

The objective of this work was to analyse and compare the prediction ability of two different approaches used to predict RFI GEBVs of genotyped selection candidates when removing their phenotypes. The first approach was the prediction of RFI GEBVs using a conventional univariate model for RFI, and the second implied a multi-trait model where GEBVs for RFI were derived using the selection index theory. The prediction abilities of both models were evaluated by two validation strategies: 30 replicates of a random group of animals, and validations across 9 different tests. The parameters used to assess the prediction ability were bias, dispersion and ratio of accuracies (Legarra & Reverter, 2018).

2 | MATERIALS AND METHODS

2.1 | Animals and data

Phenotypic data for DMI, ADG, MWT and UBF of 780 Hereford bulls were recorded in nine postweaning 70-day feed intake tests, using an automated system (GrowSafe Systems®) for DMI recording and following Beef Improvement Federation guidelines (BIF, 2016). They all came from the Kiyu test station placed in the south of Uruguay. Details are provided by Navajas et al. (2014).

The phenotypes for RFI for all animals were obtained as the residual (e) of the following lineal regression model:

$$\text{DMI} = \text{CG} + b_1 \text{ADG} + b_2 \text{MWT} + b_3 \text{UBF} + e \quad (1)$$

where CG is the feed intake test contemporary group (Test*Pen, $n=15$), DMI, ADG, MWT and UBF were previously defined, and b_1 , b_2 and b_3 are the partial regression coefficient of each trait on DMI, e is the residual of the lineal regression and represents the RFI. More details and statistical significance for each trait included in the model are described by Pravia et al. (2022a).

Additional information for WW of the animals with phenotypic RFI was included, as well as 4930 additional WW records for male calves in their contemporary groups at weaning. Three generations of pedigree (26,572 animals), as well as genomic information of 5439 animals imputed to the 50k SNP (BovineSNP50, Illumina, San Diego, CA), were included in the analysis (Table 1).

2.2 | Models

Two models were used for the estimation of GEBVs for RFI and representation matrix and description of each effect is presented below.

2.2.1 | Univariate model (Model 1)

Univariate model for RFI was implemented in BLUPF90 (Miszta et al., 2016) as follows:

$$y = Xb + Zu_{\text{RFI}} + e \quad (2)$$

where y is phenotypic RFI obtained from the linear regression model (1), X is the incidence matrix for fixed effects of age of the animal in days at the start of test, as a covariate, and age of dam class variable with five levels, Z is the matrix linking animals to phenotypes, u_{RFI} is vector of the RFI GEBV (GEBV_RFIg 1) for all animals, e is the vector of the residual effects and it was assumed that $u_{\text{RFI}} \sim N(0, H\sigma^2_u)$ and $e \sim N(0, I\sigma^2_e)$. Matrix H is the combination of pedigree-based relationships and differences between pedigree-based and genomic-based relationships (Aguilar et al., 2010).

2.2.2 | Multi-trait model (Model 2)

A five-trait model was implemented which included 780 records of animals with DMI, ADG, MWT and UBF data and WW and 4930 WW records from animals present in their contemporary groups at weaning (235 groups). Matrix representation is presented as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} X1\beta_1 \\ X2\beta_2 \\ X3\beta_3 \\ X4\beta_4 \\ X5\beta_5 \end{bmatrix} + \begin{bmatrix} Z1u_1 \\ Z2u_2 \\ Z3u_3 \\ Z4u_4 \\ Z5u_5 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} \quad (3)$$

where y_{1-5} is DMI, ADG, MWT, UBF, WW, X and Z the relation matrix described early. Genetic variances and covariances estimates used were estimated and presented by Pravia et al. (2022b) although the number of records for a multi-trait analysis were limited. Solutions

TABLE 1 Number of animals with different information.

	Number of animals
Animals with feed efficiency phenotypes (DMI, ADG, MWT, UBF, RFI)	780
Animals with WW phenotypes	5.708
Animals with genotypes	5.439
Animals with phenotypes or genotypes	10.239
Animals with genotypes and no phenotype	4.531
Sires without phenotype or genotype	16.333
Total evaluated animals	26.572

Abbreviations: ADG, average daily gain; DMI, dry matter intake; MWT, metabolic mid-test weight; RFI, residual feed intake; UBF, ultrasound back fat; WW, weaning weight.

for the predicted GEBV for the five traits were obtained, but only GEBVs for DMI, ADG, MWT and UBF were combined as in a selection index using selection index theory as follows:

$$\text{GEBV_RFI}_{g2} = \hat{u}_1 - b_2\hat{u}_2 - b_3\hat{u}_3 - b_4\hat{u}_4 \quad (4)$$

where GEBV_RFI_{g2} is the estimated GEBV for the proposed multi-trait RFI , u_{1-4} are the GEBVs for DMI (u_1), ADG (u_2), MWT (u_3) and UBF(u_4) predicted from the model (3), and the RFI index coefficients b_2 , b_3 and b_4 calculated as:

$$\mathbf{V}^{-1} = \begin{bmatrix} b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \sigma_2^2 & \sigma_{u2u3} & \sigma_{u2u4} \\ \sigma_{u2u3} & \sigma_3^2 & \sigma_{u3u4} \\ \sigma_{u2u4} & \sigma_{u3u4} & \sigma_4^2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sigma_{u1u2} \\ \sigma_{u1u3} \\ \sigma_{u1u4} \end{bmatrix} \quad (5)$$

where $\mathbf{V}$ are the matrix of genetic variances and covariances for ADG (σ_2^2), MWT (σ_3^2) and UBF (σ_4^2), $\mathbf{C}$ the genetic covariances between DMI with ADG ($\sigma_{u1,u2}$), with MWT ($\sigma_{u1,u3}$) and with UBF ($\sigma_{u1,u4}$) (Table 2), which were previously estimated in Pravia et al. (2022b) using AIREML (Misztal et al., 2016). GEBVs for both models were obtained through Single Step GBLUP (ssGBLUP) procedure implemented in BLUPF90 (Aguilar et al., 2010; Misztal et al., 2016).

2.3 | Validation strategies

Two validation strategies were performed to evaluate the predictive ability of the univariate and multivariate models.

In the Random Folds Validations, 25% of the phenotypes were randomly removed (validation set), and they were predicted based on the remaining 75% phenotypes data (training set). This procedure was replicated 30 times to obtain 30 replicates and a distribution of the prediction ability of each model.

The second strategy was the Test Folds Validation, in which phenotypes of one of the nine feed efficiency test were eliminated at each round (validation set), and GEBVs were predicted with the data of the other eight tests (training set).

In both validation strategies, for Model 1 RFI was the phenotype removed in the validation set, and their GEBVs_RFIg 1 were predicted based on Equation (2) and computed by ssGBLUP and their genomic relationship with animals with data.

In the case of Model 2, the records for DMI, ADG, MWT and UBF were removed in the validation sets, meanwhile WW was retained. The GEBV_RFI_{g2} were predicted using

Equations (3) and (4) and their genomic relationship with animals with data.

2.4 | Prediction ability statics

The predictive ability of both models was evaluated by the following parameters: bias (b_0), dispersion (b_1), ratio of accuracy (r) and relative increase in accuracy by adding phenotypic information (IP) in the focal set, as in Bermann et al. (2020) that adapted the method presented by Legarra and Reverter (2018). This method considers two data sets: the whole data set (w) which contains all phenotypes, and the partial set (p) which contains phenotypes up to a given date. The focal individuals are usually defined as the group of young animals of interest such as animals that might be selected at a given point in time given early (partial) information. This partial data set can be interpreted as the evaluation at the time of selection decision, where not information is available, and the whole data set (w) is a posteriori confirmation of the goodness of these predictions. The suggested statistics are computed as:

1. Bias (b_0): $\hat{u}_p - \hat{u}_w$, refers to the difference between the average prediction for the partial set (p) and the whole set (w), with and expected value it 0 for an unbiased prediction.
2. Dispersion (b_1): $b_{w,p} = \frac{\text{cov}(\hat{u}_w, \hat{u}_p)}{\text{var}(\hat{u}_p)}$, represent the prediction bias, for the slope of the regression of the breeding values computed as GEBVs with the whole data set ($\hat{u}_w$) on the GEBVs estimated with the partial data set ($\hat{u}_p$), with an expectation value of one if the evaluation is neither underdispersed nor overdispersed.
3. Ratio of accuracy ($\rho_{w,p}$) = $\text{cor}(\hat{u}_w, \hat{u}_p)$, defined as the sample Pearson correlation coefficient between the average prediction for the whole set (w), and the average prediction for the partial set (p), as $\hat{u}_w$ and $\hat{u}_p$, which is equal to $\frac{\text{cov}(\hat{u}_w, \hat{u}_p)}{\sqrt{\text{var}(\hat{u}_w)\text{var}(\hat{u}_p)}}$.
4. The relative increase in accuracy by adding phenotyping information in the partial set: $\text{incr}_{\text{phen}}(\text{IP}) = \rho_{w,p}^{-1} - 1$, which estimates the ratio between the accuracies obtained with the partial and whole data sets (i.e. $\text{acc}_p / \text{acc}_w$).

In this study, the "Whole population" is the complete data with all animals, genotypes and pedigree information, and the "Partial population" is the set we are interested to predict without the phenotypes (the validation set from the different validation strategies). The vector of the "Whole population value" are the GEBVs predicted with all phenotypes, genotypes and pedigree information (GEBV_RFI_w) and are used as the reference to validate with the "Partial" GEBVs obtained from the validation

sets (GEBV_RFI_p) and to compute the different statistics. Bias (b_0) is obtained as $-\hat{u}_{\text{GEBV_RFI_w}} - \hat{u}_{\text{GEBV_RFI_p}}$, dispersion (b_1) as the slope of regression of $\hat{u}_{\text{GEBV_RFI_w}}$ on $\hat{u}_{\text{GEBV_RFI_p}}$ as expressed above. The $\hat{u}_{\text{GEBV_RFI_w}}$ and $\hat{u}_{\text{GEBV_RFI_p}}$ being the best available estimation for Model 1 and Model 2, as the true breeding values (TBV) are unknown.

For each model, the ratio of accuracy (ρ) is represented with the correlation between GEBV_RFI_{1_w} and GEBV_RFI_{1_p} for Model 1, and GEBV_RFI_{2_w} and GEBV_RFI_{2_p} for Model 2, which represents the best estimation of the correlation between the TBV and the predicted $\hat{u}_{\text{GEBV_RFI}}$. The last parameter IP, proposed by Bermann et al. (2020), refers to the increase in accuracy

obtained by adding genomic information to the partial population.

3 | RESULTS

GEBVs from both models when all data available were considered (phenotypes and genotypes) were highly correlated ($r=0.96$, GEBV_RFI_{1_w} from Model 1 and GEBV_RFI_{2_w} from Model 2, Figure 1). This was presented also in a previous work (Pravia et al., 2022b).

Statistics for the prediction ability for the univariate RFI Model (Model 1) are presented in Table 3 for the Random validation strategy and in Table 4 for the Test validation strategy. In Table 3, the mean, standard deviation, minimum and maximum for the 30 replicates of estimated parameters: bias (b_0), dispersion (b_1), ratio of accuracy ($\rho_{w,p}$) and increase in accuracy by adding phenotypes to the focal data set (IP) was presented. For the Test validation strategy, the estimated parameters for each test are presented in Table 4.

3.1 | Univariate model (Model 1)

The results for this analysis indicate that predictions from Model 1 were unbiased, as the b_0 parameter was 0 (± 0.01) for both the random validation strategy and across the 9 tests. The coefficient of dispersion was 0.956 (± 0.10) for the random strategy and between 0.67 and 1.17 for the test validation, which suggest that predictions were neither underdispersed nor overdispersed. Additionally, the ratio between accuracies obtained with partial and whole data set was estimated to be 0.587 (± 0.04) and 0.52 for random and test validations respectively, indicating that predictions based solely on genomic information was mostly accurate. The estimation of the IP showed that including phenotypes in Model 1 would increase the accuracy of prediction by an average of 71%.

Predictions across tests were somewhat variable. Tests 1, 7, 8 and 9 had good predictions, while test 2 and test 5 had lower predictions of GEBVs. The lowest values for b_1 and ratio of accuracy was observed in test 5, which consisted in a group of bulls from higher proportion of

TABLE 2 Genetic co-variance components and genetic correlations used for multi-trait model used in RFI₂.

	WW (kg)	DMI (kg)	ADG (kg/day)	MWT (kg)	UBF (mm)
WW	249.6	6.65	1.47	72.46	3.24
DMI	0.54	0.61	0.11	4.63	0.24
ADG	0.57	0.83	0.03	0.88	0.04
MWT	0.66	0.85	0.78	48.43	2.22
UBF	0.40	0.61	0.43	0.62	0.26

Note: Additive genetic variance in diagonal (in bold), genetic covariances above and genetic correlations below the diagonal.

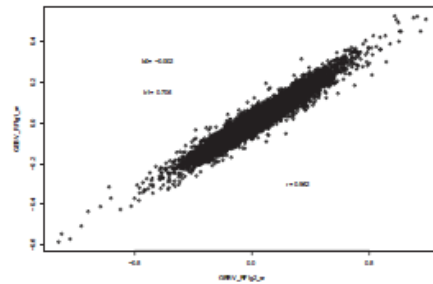


FIGURE 1 Scatter plot between GEBVs obtained for the RFI univariate model (Model 1) and the GEBVs obtained from the multivariate model (Model 2) for all animals with complete data.

TABLE 3 Prediction ability indicators for RFI_univariate (Model 1) for the 30 random replicates.

Model 1 – RFI_univariate	Mean	SD	Min	Max
Bias (b_0)	0.003	0.008	-0.018	0.020
Dispersion (b_1)	0.956	0.100	0.790	1.130
Ratio of accuracy ($\rho_{w,p} = \text{cor}(\hat{u}_w, \hat{u}_p)$)	0.587	0.044	0.480	0.670
incr _{phen} (IP) = $\rho_{w,p}^{-1} - 1$	0.710	0.130	0.490	1.080

TABLE 4 Validations coefficients for RFI_univariate (Model 1) across 9 Tests.

Model 1	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Avg
- RFI_univariate										
Bulls ^a	49	11.9	49	124	44	115	103	87	90	-0.002
Bias (b0)	-0.018 (0.02)	0.008 (0.01)	0.001 (0.02)	-0.002 (0.01)	-0.018 (0.02)	-0.011 (0.01)	-0.006 (0.015)	0.006 (0.011)	-0.001 (0.01)	
Dispersion (b1)	1.074 (0.17)	0.940 (0.13)	0.802 (0.17)	1.138 (0.13)	0.647 (0.20)	0.989 (0.11)	1.170 (0.12)	0.988 (0.12)	0.980 (0.13)	1.006
Ratio of accuracy ($\rho_{w,p} = \text{corr}(\hat{a}_w, \hat{a}_p)$)	0.67	0.55	0.57	0.62	0.45	0.64	0.69	0.67	0.63	0.62
Incr _{pbm} (IP) = $\rho_{w,p}^{-1} - 1$	0.49	0.82	0.75	0.61	1.22	0.56	0.45	0.49	0.58	0.61

^aNumber of bulls considered in each validation set, 0 mean standard deviations

national sires than the other tests. However, when connectedness of this group with the rest of animals were tested by Principal Component Analysis (PCA), no notable differences were observed in comparison with the other tests, which might explain this lower performance. The statistics for prediction ability for Multivariate Model (Model 2) are presented in Table 5 for Random validation strategy and in Table 6 for Test validation strategy.

3.2 | Multivariate model (Model 2)

For these comparisons, the vector of the "Whole population value" was the GEBV_RFIg 2 predicted with all phenotypes, genotypes and pedigree information (GEBV_RFIg 2_w) obtained from combining the individual GEBVs (3) in an index (4). It comprised information of WW, DMI, ADG, MWT and UBF for the 780 animals with feed efficiency data, as well as the WW of their 4930 contemporary animals. The prediction on each partial set to validate in this model (GEBV_RFIg 2_p) was based on their individual GBVs combined also as in (3) and (4), although in this case GEBVs were predicted using only WW phenotype, genetic correlations between WW and the other traits, genomic information and genomic and pedigree relationship with all animals with phenotypic data.

Predictions by Model 2 were unbiased since the b0 estimator was 0.01 (± 0.01) for both random replicates and Test strategies. The coefficient of dispersion was 0.89 for random replicates, indicating some overdispersion of GEBVs based on partial data but also an underdispersion of GEBV based on whole data. The ratio of accuracy of predictions was lower than for Model 1, with an average of 0.51 (0.44–0.62) for the 30 replicates and 0.52 for the test validations with a range from 0.31 for test 7 to 0.59 for test 13. The indicator IP shows that the inclusion of phenotypes impact on increasing the accuracy on average in 96%.

The prediction ability across different validations strategies showed consistent results, indicating Model 1 performs better than Model 2 both in 30 random validations and test validations, as evidenced by the superior values of all prediction ability parameters (b0, b1, ratio of accuracy) for Model 1. *T*-student analysis demonstrated that b1 and $\rho_{w,p}$ were statistically different between Model 1 and Model 2, while b0 was not statistically different between models.

Predictions also showed variability across replicates or were dependent on the group of animals being predicted. In Model 1, prediction accuracy fluctuated between 0.48 and 0.67 according to the group of animals (Table 3) or from 0.45 to 0.67 among tests (Table 4). Similarly, prediction accuracies ranged from 0.44 to 0.62 across the

TABLE 5 Prediction ability indicators for RFI_multivariate (Model 2) for the 30 random replicates.

Model 2 RFI_multivariate	Mean 30 replicates	SD	Min	Max
Bias (b_0)	0.006	0.010	-0.019	0.037
Dispersion (b_1)	0.887	0.110	0.714	1.138
Ratio of accuracy ($\rho_{w,p} = \text{cor}(\hat{u}_w, \hat{u}_p)$)	0.510	0.050	0.440	0.617
$\text{incr}_{\text{phen}} = \rho_{w,p}^{-1} - 1$	0.960	0.180	0.620	1.250

30 replicates and between 0.31 and 0.63 among test with Model 2 (Tables 5 and 6). An ANOVA analysis were implemented to analyse the prediction thought the different test. Results indicates that prediction coefficients differ statically between tests also within the same model (Model 1 or 2). When we compared both models, Model 1 versus Model 2, for the same test (Tables 4 and 6), the coefficients for the same test were either equal or statistically different (T -student test), depending on the test. These results indicate that the group of animals being predicted is more relevant than the model was used, highlighting the importance of the linkage between the animals being predicted and the training population (see Table 7).

4 | DISCUSSION

Genetic (genomic) evaluations have the main purpose to accurately predict breeding values (GEBV) for all the candidates for selection. For that, the mixed model equations and BLUP programs uses different types of information, phenotypes, pedigree relationships and genotypes (Aguilar et al., 2010). When the aim is to improve accuracy of prediction for traits that are scarcely recorded, such as RFI, Calus and Veerkamp (2011) recommend different strategies with regard to the selection candidates: (1) using pedigree indexes for the indicator trait and/or the trait of interest, (2) recording the performance of an indicator trait in common sib or progeny testing schemes, (3) recording performances for the trait of interest, (4) obtaining genotypes and (5) using various combinations of these strategies. Their results demonstrate the most effective strategy depends on the genetic correlation with the indicator traits.

In this study, we explored the prediction ability of two models and the impact of information on indicator traits for relatives and contemporary animals in a multi-trait model (Model 2) to predict GEBV for RFI, in contrast to the univariate model (Model 1) for RFI. For feed efficiency traits which are very expensive to measure, the implementation of genomic prediction is a crucial tool to effectively accelerate the genetic progress in the population as only few animals can be measured. For that reason, the main objectives of this study were to investigate the prediction

ability of two alternatives models proposed in the context of genomic prediction. Both proposed models were demonstrated to give equivalent GEBV for RFI, with a correlations of 0.96 (Figure 1). However, it was necessary to evaluate their performance in the absence of phenotypes where their GEBV were predicted based solely on genotypes.

With the increase of using genomic information for livestock animals, validation techniques became widely used in animal breeding, especially to validate genomic models (Gianola & Schon, 2016). Nevertheless, the discussion for the statistical validation of genetic models for genetic evaluation has been always a topic of concern (Thompson, 2001, Legarra & Reverter, 2018). Usually, three measures are taken to evaluate the quality of genetic predictions: accuracy, bias and dispersion. Legarra and Reverter (2018) pointed out that it is necessary to have tools to rank, understand and quantify the behaviour of prediction models to decided which model is the best. According them, theoretically best predictors, defined as conditional expectations, have optimal selection proprieties, and therefore we aim got models that yield such best predictors. These authors proposed an approach to test the quality of evaluation methods using cross. validation test based on successive evaluations of EBVs of a set of "partial" o "focal" individuals of interest, and a "whole" data which is based on the use of "old" and "recent + old" and they proposed of a series of statics to evaluate the prediction accuracy.

These statistics were used in our study to evaluate the prediction ability of both models proposed. We had to make some modifications since some of the premises of LR method did not apply to the available data. We do not have many generations of data to evaluate the successive partition of data in "old" or "recent" groups of animals. As there are few years of accumulated data and GEBVs for RFI were recently published no clear selection has been possible yet. For these reasons and to get more replicates, we adapted the method to randomly extract 25% of data and repeated these predictions 30 times and to analyse the distribution of predictions. Although we used a random sampling of 30 samples, each animal had the chance to be present in multiple samples, thus the folds were not independent. This lack of independence among samples

TABLE 6 Validations coefficients for RFI_multivariate (Model 2) across 9 Tests.

Model 2	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Avg
RFI_multivariate										
Bulls ^a	49	119	49	124	44	115	103	87	90	
Bias (b0)	0.002 (0.03)	0.001 (0.02)	-0.01 (0.03)	-0.006 (0.02)	0.007 (0.03)	-0.005 (0.02)	-0.009 (0.02)	0.00 (0.02)	-0.003 (0.02)	-0.023
Dispersion (b1)	0.961 (0.20)	0.923 (0.15)	0.684 (0.19)	0.993 (0.15)	0.471 (0.23)	0.987 (0.13)	1.107 (0.13)	0.88 (0.13)	0.908 (0.15)	0.88
Ratio of accuracy ($\rho_{w,p} = \text{cor}(\hat{a}_w, \hat{a}_p)$)	0.58	0.49	0.46	0.51	0.31	0.57	0.63	0.59	0.54	0.52
$\text{incr}_{\text{phen}} = \rho_{w,p}^{-1} - 1$	0.72	1.04	1.15	0.97	2.24	0.74	0.58	0.68	0.85	0.998

^aNumber of bulls considered in each validation set, () mean standard deviations

could lead to non-independent means and may affected the accuracy of our findings. Perhaps the consideration of an alternative strategy, such as the “leave one out”, could avoid this problem. However, when we performed the evaluation using different tests, in which the samples are independent of each other as they were obtained from different animals, the results were consistent. This suggests that the predictions generated by these models exhibit similar levels of precision, ranging between 0.40 and 0.67, regardless of the validation strategy used.

The estimators used in this study to evaluate the predictive ability were bias, dispersion and ratio of accuracy as described by Legarra and Reverter (2018). The results of this study demonstrated that both Model 1 and Model 2 had good predictive ability, allowing accurate predictions of GEBV for not phenotyped animals based on their genotype. Both models were unbiased, as evidenced by the parameter b0, which was zero when evaluated by both validation strategies. This means that the GEBV prediction mean for the partial set analysed (p) were the same as for the whole dataset (w), as we expected and pointed out in Legarra and Reverter (2018). Bias of genetic predictions have been a concern topic on genomic prediction (Macedo et al., 2020), particularly for traits that have been genomically preselected. In our study as selection has not been in place, no bias was observed. The next parameter we considered to evaluate the predictive ability was the dispersion (b1), interpreted as the slope of the regression of GEBV_RFI_w on GEBV_RFI_p, and is estimated as $b_{wp} = \text{cov}(\hat{a}_w, \hat{a}_p) / \text{var}(\hat{a}_p)$, where cov and var denote the sample and complete variance and covariance, respectively. The value b_{wp} indicate overdispersion of GEBVs based on partial data but also an underdispersion of GEBV based on whole data (Legarra & Reverter, 2018). In Model 1, the average b_{wp} of the 30 replicates was closer to 1 (0.956 for the random replicates and 1.00 for the test validation), while for Model 2, the average b_{wp} of all replicates and test validations was 0.88, lower than the value obtained for Model 1. This indicates that predictions based on Model 2 could be slightly overestimated.

The last indicator was the ratio of accuracy, theoretically defined as the sample Pearson correlation coefficient between the TBV (not known), and the estimated breeding value (GEBV) represented as $\hat{a}_w$ and $\hat{a}_p$, respectively. This ratio of accuracy can be used as an estimation of accuracy and is equal to $\text{cov}(\hat{a}_w, \hat{a}_p) / \sqrt{\text{var}(\hat{a}_w) \text{var}(\hat{a}_p)}$, which Legarra and Reverter (2018) refers to as “predictivity”. The differences between models were more evident in this parameter. The mean ratio of accuracies obtained for Model 1 was 0.587 for random replicates and 0.62 for test validation strategy, while it was 0.51 and 0.52, respectively, for Model 2. These results indicate a higher association between the GEBVs considered with “whole” data (w) as

TABLE 7 Acronyms.

RFI	Phenotypic residual feed intake, is the residual of the lineal regression of ADG, MWT and UBF on DMI
$\hat{u}_{RFI}$	Vector of solutions for the RFI GEBV obtained from Univariate Model 1
GEBV_RFIg 1	Refers to the GEBVs obtained from the univariate model (Model 1) in the explanation from the general model independently which population is considered
GEBV_RFIg 2	Refers to the GEBVs obtained from the multi-trait model (Model 2) in the explanation from the general model independently which population is considered
GEBV_RFIg 1_w	GEBVs for all animals obtained from Model 1 predicted from all phenotypes, genotypes and pedigree information available
GEBV_RFIg 1_p	GEBVs for the partial group of candidates animals obtained from Model 1 predicted from genotypes and pedigree information, without phenotypes
GEBV_RFIg 2_w	GEBVs for all animals obtained from Model 2 predicted from all phenotypes, genotypes and pedigree information available
GEBV_RFIg 2_p	GEBVs for the partial group of candidates animals obtained from Model 2 predicted from genotypes and pedigree information, without phenotypes

the closest approximation to the TBV and the sGEBVs predicted with the partial data (*p*). The increase in accuracy by adding phenotypic information (IP) was 71% for Model 1 and 96% for Model 2, which also reflect the more accurate prediction for Model 1.

The results also showed that estimated coefficients across the 30 replicates or tests revealed some dispersion between them. This could be consequence of the group of animals included in each partial set to be predicted set (*p*) in relation with the “whole” data (*w*). Pszczola, Strabel, Mulder, and Calus (2012), Pszczola, Strabel, van Arendonk, and Calus (2012) studied the effect of the relationship between the animals on the reference population and the set to be predicted, especially when the reference population consisted of a limited number of individuals. In that situation, the relationship should be maximized, until a big reference population could be achieved. In our situation, with small data available, prediction is highly dependent on the relationships. However, principal component analysis showed no significant distances between different tests and groups of animals.

As discussed above, the results of this study indicate that Model 1 performs better than Model 2 in predicting non phenotyped animals based on genotype information. These findings are surprising given previous results from Pravia et al. (2022b), which Model 2 had higher individual GEBV accuracies due to its use of information from highly correlated traits. Nevertheless, this study did not confirm this expectation.

Several authors have explored the potential of using predictor traits in a multi-trait analysis as an inexpensive way to predicted GEBVs and improve individual accuracy for trait that are difficult to measure, such as DMI and RFI (MacNeil et al., 2011; Manzanilla-Pech et al., 2017; Pszczola et al., 2013). For instance, MacNeil et al. (2011), Pszczola et al. (2013) and Manzanilla-Pech

et al. (2017) have investigated this approach. In a study by MacNeil et al. (2011), the authors integrated WW into a multi-trait genetic evaluation to predict EBVs for RFI in Angus cattle; however, they did not conduct a validation analysis to evaluate the prediction accuracy. Other researchers have reported the advantages of using multi-trait genomic models over univariate models to achieve more accurate and less biased GEBVs (Calus & Veerkamp, 2011; Guo et al., 2014; Jibrila et al., 2021; Pszczola et al., 2013). Manzanilla-Pech et al. (2017) proposed the inclusion of predictor traits in dairy cattle to predict DMI, but no benefit was observed by adding information on predictor traits to the reference population when DMI was already included. However, in the absence of DMI records, records on proxy traits were a good way to obtain EBVs with relatively good accuracy. These authors explained that the highest accuracy of DMI prediction would be determined by the magnitude of the genetic correlation between the proxy and the trait of interest. With the high genetic correlation observed between predictor traits and indicator, it is possible to estimate EBV for DMI by using predictor traits, such as fat and protein corrected milk (FPCM) and live weight (LW). They also found when considering only the predictor traits in the reference population, the accuracies of estimated EBV for DMI were lower than in the scenarios using DMI. When the predictor trait was added when DMI were already included, there were no or slight differences in accuracies. Pszczola et al. (2013) also found recording predictor traits for both the reference population and evaluated animals significantly increased DMI breeding value accuracy and removed the bias observed when only references animals had records. Calus and Veerkamp (2011) develop three different models that permit multi-trait genomic selection by combining scarcely records trait with genetically correlated indicator traits

and to compare their performance to single-trait models. They found that according to the level of genetic correlation between the trait of interest and the indicator traits belong genotypes in selection candidates is more effective than obtaining phenotypes for an indicator trait. When genetic correlation was high, above 0.5, using an indicator trait led to higher accuracies for selection candidates than genotyping the individual. However, if the correlation is below that, the accuracy was higher when the candidate was genotyped rather than obtaining phenotypes for the indicator trait. Based on the high genetic correlation between WW and DMI, ADG, MWT and UBF, which were 0.54, 0.57, 0.66 and 0.40, respectively (Table 2), we expected WW to be a good proxy for these traits, leading to better prediction of their EBVs and improved performance of calculated GEBV_RFIg 2 from selection indexes (4 and 5). However, the inclusion of WW phenotypes in the multi-trait model (Model 2) to predict the GEBV_RFIg 2 did not provide improved information compared to the genotypes already considered. Predictions were also lower than those obtained directly through the genotypes in the univariate model. One possible explanation could be that although WW is a good proxy trait for DMI, ADG, MWT, it might not be an accurate predictor of RFI. When we tested the ratio of accuracy when predicting GEBV_RFIg 2 with or without the phenotype of WW in the validation set showed that including WW in the model (3) resulted in better prediction ability (0.514 vs. 0.502, results not showed). This is likely due to the better estimation of individual GEBVs resulting from the solutions of the multi-trait model (Model 3) and the combined index (Model 4), as WW improved the prediction of each individual GEBV for all traits. However, it is possible that WW could introduce additional variation that does not explain individual and biological variation for RFI, which could negatively impact the prediction. To confirm the independence of RFI from WW, we estimated a genetic correlation between WW and RFI of -0.06 ± 0.23 . This suggests that WW and RFI are genetically independent, but the estimate had a large standard deviation because of the dataset size.

While this study focused on a population where all animals to be predicted were genotyped, in practice it is not feasible to genotype all potential candidates. In such situations, the GEBVs for animals that are not phenotyped or genotyped can be estimated using the H relationship matrix (Aguilar et al., 2010) and predictions will be based on additive genetic relationship with the training population. The predictability for those animals will depend on how well connected they are through the A matrix to the training population. In that situation the use of correlated traits in non-genotyped animals could be a promising way to improve prediction and should be studied in further analysis.

5 | CONCLUSIONS

This study shows it is possible to predict accurate and unbiased RFI GEBVs for non phenotyped selection candidates based on genomic prediction. The results of both validation strategies indicated that the univariate model performed better than the multivariate model. Even when the multivariate model included a predictor trait, its predictive ability was not superior to that of the univariate model when all focal animals to be predicted were genotyped. In a situation where candidates to be predicted are not genotyped, the use of a multi-trait model including a correlated trait could potentially conduct to better results, leading to different outcomes compared with our current findings. Further studies should be conducted to investigate this possibility.

FUNDING INFORMATION

None.

CONFLICT OF INTEREST STATEMENT

The authors declare that they do not have any conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Maria Isabel Pravia  <https://orcid.org/0000-0003-1576-4495>

Elly Ana Navajas  <https://orcid.org/0000-0002-9018-4741>

Ignacio Aguilar  <https://orcid.org/0000-0002-1038-4752>

Olga Ravagnolo  <https://orcid.org/0000-0002-3253-095X>

REFERENCES

- 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 score1. *Journal of Dairy Science*, 93(2), 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Basarab, J. A., Price, M. A., Aalhus, J. L., Okine, E. K., Snelling, W. M., & Lyle, K. L. (2003). Residual feed intake and body composition in young growing cattle. *Canadian Journal of Animal Science*, 83, 189–204.
- Bermann, M., Legarra, A., Hollifield, M. K., Masuda, Y., Lourenco, D., & Misztal, I. (2020). Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. *The Journal of Animal Breeding and Genetics*, 138, 4–13.

- BIF. (2016). Guidelines for Uniform Beef Improvement Programs: beef Improvement Federation Recommendation.
- Calus, M., & Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetic Selection Evolution*, 43, 26. <https://doi.org/10.1186/1297-9686-43-26>
- Gianola, D., & Schon, C. C. (2016). Cross-validation without doing cross-validation in genome-enabled prediction. *G3-Genes Genomes Genetics*, 6(10), 3107–3128. <https://doi.org/10.1534/g3.116.033381>
- Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., & Guosheng, S. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genetics*, 15, 30. <https://doi.org/10.1186/1471-2156-15-30>
- Hayes, B., Bowman, P., Chamberlain, S., & Goddard, M. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 91, 433–443. <https://doi.org/10.3168/jds.2008-1646>
- Jibrila, I., Vandenplas, J., ten Napel, J., Veerkamp, R. F., & Calus, M. P. L. (2021). Avoiding preselection bias in subsequent single-step genomic BLUP evaluations of genomically preselected animals. *The Journal of Animal Breeding and Genetics*, 138, 432–441.
- Kennedy, B. W., Van der Welf, J. H. J., & Meuwissen, T. H. E. (1993). Genetic and statistical properties of residual feed intake. *The Journal of Animal Science*, 71, 3239–3250. <https://doi.org/10.2527/1993.71123239x>
- Legarra, A., & Reverter, A. (2017). Can we frame and understand cross-validation results in animal breeding? *Proceedings of the Association for the Advancement of Animal Breeding and Genetics*, 22, 73–80.
- Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, 50(1), 1–18. <https://doi.org/10.1186/s12711-018-0426-6>
- Macedo, F. L., Reverter, A., & Legarra, A. (2020). Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *Journal of Dairy Science*, 103, 529–544. <https://doi.org/10.3168/jds.2019-16603>
- MacNeil, M. D., Lopez-Villalobos, N., & Northcutt, S. L. (2011). A prototype national cattle evaluation for feed intake and efficiency of Angus cattle. *Journal of Animal Science*, 89, 3917–3923. <https://doi.org/10.2527/jas.2011-4124>
- Mäntysaari, E. A., Liu, Z., & VanRaden, P. (2010). Interbull validation test for genomic evaluations. *Interbull Bulletin*, 41, 17–21.
- Manzanilla-Pech, C. I. V., Veerkamp, R. F., de Haas, Y., Calus, M. P. L., & ten Napel, J. (2017). Accuracies of breeding values for dry matter intake using nongenotyped animals and predictor traits in different lactations. *Journal of Dairy Science*, 100, 9103–9114. <https://doi.org/10.3168/jds.2017-12741>
- Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., & Vitezica, Z. (2016). *Manual for BLUPF90 family of programs*. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf
- Navajas, E. A., Pravia, M. I., Lema, M., Clariget, J., Aguilar, I., Ravagnolo, O., Brito, G., Peraza, P., Dalla Rizza, M., & Montossi, F. (2014). Genetic improvement of feed efficiency and carcass and meat quality of Hereford cattle by genomics. *Archivos Latinoamericanos de Producción Animal*, 22(5), 13–16.
- Pravia, M. I., Navajas, E. A., Aguilar, I., & Ravagnolo, O. (2022a). Evaluation of feed efficiency traits in different Hereford populations and their effect on variance component estimation. *Animal Production Science*, 62(17), 1652–1660. <https://doi.org/10.1071/AN21420>
- Pravia, M. I., Navajas, E. A., Aguilar, I., & Ravagnolo, O. (2022b). *Alternative models to predict residual feed intake in Hereford breed and effects on their breeding values accuracy*. *World Congress on Genetics Applied to Livestock Production (WCGALP)*, 12. https://doi.org/10.3920/978-90-8686-940-4_48
- Pszczola, M., Strabel, T., Mulder, H. A., & Calus, M. P. L. (2012). Reliability of direct genomic values for animals with different relationships with and to the reference population. *Journal of Dairy Science*, 95, 389–400. <https://doi.org/10.3168/jds.2011-4338>
- Pszczola, M., Strabel, T., van Arendonk, J. A. M., & Calus, M. P. L. (2012). The impact of genotyping different groups of animals on accuracy when moving from traditional to genomic selection. *Journal of Dairy Science*, 95, 5412–5421. <https://doi.org/10.3168/jds.2012-5550>
- Pszczola, M., Veerkamp, R. F., de Haas, Y., Wall, E., Strabel, T., & Calus, M. P. L. (2013). Effect of predictor traits on accuracy of genomic breeding values for feed intake based on limited cow references population. *Animal*, 7(11), 1759–1768. <https://doi.org/10.1017/S175173111300150X>
- Ravagnolo, O., Aguilar, I., Crowley, J. J., Pravia, M. I., Lema, M., Macedo, F. L., Scott, S., & Navajas, E. A. (2018). *Accuracy of genomic predictions of residual feed intake in Hereford with Uruguayan and Canadian training populations*. Abstract, WCGALP, 2018.
- Reverter, A., Golden, B. L., Bourdon, R. M., & Brinks, J. S. (1994). Technical note: detection of bias in genetic predictions. *Journal of Animal Sciences*, 72(1), 34–37. <https://doi.org/10.2527/1994.72134x>
- Saatchi, M., McClure, M. C., McKay, S. D., Rolf, M. M., Kim, J. W., Decker, J. E., Taxis, T. M., Chapple, R. H., Ramey, H. R., Northcutt, S. L., Bauck, S., Woodward, B., Dekkers, J. C. M., Fernando, R. L., Schnabel, R. D., Garrick, D. J., & Taylor, J. F. (2011). Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution*, 43, 40. <https://doi.org/10.1186/1297-9686-43-40>
- Thompson, R. (2001). Statistical validation of genetic models. *Livestock Production Science*, 72(1–2), 129–134.
- VanRaden, P. M., VanTassel, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for north American Holstein bulls. *Journal of Dairy Science*, 92, 16–24. <https://doi.org/10.3168/jds.2008-1514>

How to cite this article: Pravia, M. I., Navajas, E. A., Aguilar, I., & Ravagnolo, O. (2023). Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake. *Journal of Animal Breeding and Genetics*, 00, 1–11. <https://doi.org/10.1111/jbg.12775>

6. Discusión general

La eficiencia de conversión de alimento tiene un alto impacto en los sistemas de producción de carne, ya que los costos de alimentación representan más del 75 % de los costos. La inclusión de esta característica en los programas de mejora genética ha sido limitada por la dificultad de medir el consumo individual en condiciones comerciales, lo que lleva a que solo una baja proporción de los candidatos a la selección dispongan de estimación de mérito genético para esta característica basada en el registro fenotípico propio.

La incorporación de la selección genómica para la mejora genética en este tipo de características abre otras posibilidades para ser aplicadas a nivel comercial. En Uruguay se llevó a cabo un proyecto que permitió recabar registros de consumo y desempeño productivo de toros y novillos Hereford provenientes de cabañas distribuidas en todo el país con el fin de construir una población de referencia o entrenamiento para aplicar selección genómica para eficiencia de conversión.

Contando con la información recabada de los cuatro años del proyecto fue que se planteó este trabajo de tesis. En primer objetivo fue estimar parámetros genéticos nacionales para esta característica; posteriormente, además de analizar los modelos convencionales, se propuso plantear otros modelos alternativos que *a priori* presentaban ventajas en términos de las precisiones en las estimaciones de los valores de cría y en la habilidad predictiva. Para ello se realizaron estudios de validación utilizando distintas alternativas en la partición de datos y se analizaron los indicadores correspondientes.

En cuanto a las implicancias de este trabajo, el objetivo fue abordar estos temas con el fin de generar información científica nacional valiosa para esta nueva característica. Se propusieron y discutieron aspectos metodológicos con el fin de generar antecedentes que podrán ser considerados en futuras aplicaciones en los programas de mejora genética en bovinos de carne.

A continuación, se resumen y se discuten las principales conclusiones de los distintos trabajos que fueron abordados específicamente en cada uno de los artículos científicos publicados.

6.1. Estimación de parámetros genéticos para eficiencia de conversión

Conocer los parámetros genéticos de una nueva característica para ser incorporada en un programa de mejora genética es crucial a la hora de determinar el potencial de respuesta a la selección. Por esa razón, el primer objetivo de este trabajo fue el estudio de la variabilidad genética y heredabilidad de consumo de alimento, consumo residual de alimento y sus regresores para la población Hereford del Uruguay.

En el caso de este tipo de características de costosa medición, la principal limitante para estimar parámetros genéticos es contar con suficiente número de fenotipos medidos a escala comercial. Por ese motivo, numerosos acuerdos internacionales de colaboración entre países y bases de datos se han venido desarrollando con el fin de incrementar las poblaciones de referencia y expandir los resultados (Berry et al., 2014). Con este mismo sentido fue que para este trabajo, además de la información recabada a nivel nacional, se contó con una base de datos canadienses a los efectos de incrementar el número de fenotipos y estimar parámetros de forma más precisa.

Los datos nacionales consistieron en registros de consumo de toros y novillos Hereford medidos en la estación de prueba de la central de Kiyú. Los datos provenientes de novillos fueron descartados para la propuesta estudiada en esta tesis, ya que una muy baja proporción de los animales contaba con genealogía completa, por lo cual no aportaban información a la estimación de parámetros genéticos. En cuanto a la información proveniente de las nueve pruebas de toros, los 780 animales provenían de 78 cabañas distribuidas en todo el país, las cuales estaban conectadas a la población Hereford nacional. Con respecto a los datos provenientes de Canadá, se contó con 1597 registros de consumo y variables de animales Hereford medidos en pruebas en dicho país.

En primer lugar, se compararon fenotípicamente ambas bases de datos a los efectos de evaluar si las características se podían considerar las mismas para la población de Uruguay y Canadá y si eran posibles de ser analizadas juntas o si se trataba de diferentes variables. Para ello se realizaron estudios descriptivos y

comparaciones estadísticas entre ellas. Los resultados fueron publicados en el primer artículo (Pravia et al., 2022a). Dichos resultados mostraron algunas diferencias fenotípicas, especialmente para peso vivo y espesor de grasa, entre los dos grupos de animales considerados, siendo mayores para los animales canadienses.

Con respecto a la primera hipótesis planteada en la tesis, los resultados demostraron que las distintas poblaciones estudiadas, Uruguay y Canadá, presentan heterogeneidad de componentes de varianzas y correlaciones genéticas entre las características. Las estimaciones realizadas para Canadá fueron más precisas, lo que era de esperar al disponer de un mayor número de registros, mientras que, para Uruguay, las estimaciones se vieron afectadas con mayores errores estándar. Esto dificultó en alguna medida la comparación de parámetros entre países. En ese sentido, se buscó evaluar a nivel de correlaciones genéticas si para ambas poblaciones podrían tratarse como la misma característica. Planteamos modelos bivariados para cada característica (DMI, ADG, PMET, BFAT y RFI), considerando en cada país la característica, con el fin de estimar la correlación genética entre países. Sin embargo, los datos no fueron lo suficientemente contundentes para poder estimar dicha correlación: las soluciones de AIREML no fueron estimables y los estudios de correlaciones realizados a través de Gibbs *sampling* dieron valores muy dispersos. A pesar de esto, y basándonos en antecedentes internacionales que integran información proveniente de diversos orígenes, concluimos que la combinación de ambas bases de datos sería la mejor opción para estimar los parámetros de manera más precisa, considerando el país y el grupo contemporáneo de medición como efectos en el modelo. Esto permitió obtener un mayor número de fenotipos, lo que se tradujo en estimaciones más precisas para la variabilidad genética y heredabilidad para estas características.

Como primer resultado de este trabajo, se estimaron los parámetros genéticos, los cuales fueron coincidentes con lo reportado por otros autores para estas características en animales Hereford (Herd y Bishop, 2000) y Angus (Arthur et al., 2001). El consumo de alimento presentó una heredabilidad de 0,42 ($\pm 0,12$) en la población de URY, h^2 de 0,41 ($\pm 0,07$) para CAN y 0,45 ($\pm 0,07$) para la población conjunta. Estos resultados indican que, al seleccionar por un menor consumo, es

posible esperar respuestas positivas a la selección debido a la alta heredabilidad de esta característica. Con respecto al RFI, se pudo estimar alguna diferencia entre las poblaciones, con una heredabilidad de 0,34 ($\pm 0,13$) para la población de URY, 0,20 ($\pm 0,06$) para CAN y 0,25 ($\pm 0,06$) para la población conjunta.

Los resultados obtenidos en este trabajo permitieron corroborar la segunda hipótesis planteada. El consumo de alimento resultó estar positivamente correlacionado con ganancia, peso y espesor de grasa (0,68, 0,70 y 0,39, respectivamente, para el set de datos conjunto). Esto tiene importantes implicancias en el uso de estas características como criterio de selección, ya que significa que la selección por un menor consumo va a traer aparejados cambios desfavorables en el peso, ganancia diaria de peso y el espesor de grasa. Por el contrario, la selección por el RFI no parecería tener consecuencias sobre las demás características, ya que las correlaciones genéticas con sus regresores fueron entre -0,12 y 0,07, estadísticamente no diferentes a 0.

En conclusión, y coincidiendo con lo reportado en la bibliografía, los resultados en este trabajo permiten concluir que tanto el consumo como el RFI presentan variabilidad genética y heredabilidades medias. Desde una perspectiva de herramienta de selección, el uso de los valores de cría o DEP para RFI permitiría reducir el consumo sin afectar el desempeño productivo, lo cual presentaría ventajas con respecto al uso de consumo como criterio de selección, ya que esta última tendría consecuencias indeseables sobre otras características productivas de relevancia económica.

6.2. Evaluación de modelos alternativos para predecir valores de cría para RFI

La forma convencional de considerar la característica RFI en las evaluaciones genéticas requiere de dos pasos. En primer lugar, se predice el consumo a partir de una regresión lineal en función del peso, ganancia y espesor de grasa y se obtiene el residuo (RFI) como la diferencia entre el consumo real observado y el consumo predicho a partir de los predictores regresores (ecuación 1 de la revisión bibliográfica). Este residuo se utiliza como el fenotipo en el modelo genético (ecuación 2), el cual, mediante la metodología BLUP, permite obtener los DEP para

esta característica tanto para aquellos animales con el fenotipo como para sus parientes cercanos. En el caso de una población genotipada, las predicciones genómicas se extienden a aquellos individuos que no tienen fenotipo, pero que sí disponen de genotipo, haciendo uso de las ventajas del método SSGBLUP (Aguilar et al., 2010).

Sin embargo, esta consideración del RFI puede prestar algunas restricciones que han sido discutidas por algunos autores (Crews, 2005; Kennedy et al., 1993; Lu et al., 2015; Tempelman et al., 2015).

1. El RFI por definición, y de acuerdo a cómo es calculado, es independiente fenotípicamente de los términos de regresión *energy sinks* (ganancia diaria, peso metabólico, espesor de grasa) (Arthur et al., 2001; Basarab et al., 2003). Sin embargo, puede no ser genéticamente independiente de los regresores del modelo, lo que tendría fuertes implicancias en las respuestas a la selección (Kennedy et al., 1993; Crews, 2005).

2. El RFI, al ser un indicador no observable, puede ser más difícil de interpretar y no presenta ninguna ventaja frente al RFI derivado directamente de sus regresores. Fue demostrado matemáticamente que el RFI puede ser obtenido directamente a partir de regresiones genéticas entre el consumo y las características predictoras; también que es equivalente al propuesto por regresión lineal (Kennedy et al., 1993; Lu et al., 2015) y es más comprensible por los usuarios (VanderWerf, 2004).

3. Animales que participan de las pruebas habitualmente son preseleccionados, lo que podría conducir a un sesgo de la estimación de los DEP para estos individuos (Pollak et al., 1984).

4. La inclusión de características indicadoras en modelos multivariados y el uso de correlaciones genéticas para corregir el sesgo ha sido una práctica ampliamente utilizada para mejorar las predicciones y expandirlas a un mayor número de animales (Crews et al., 2010; MacNeil et al., 2011; Pszczola et al., 2013; Lu et al., 2015).

5. Para este tipo de características, solo unos pocos individuos son fenotipados, por lo que una baja proporción de los candidatos cuentan con DEP estimados de forma precisa y al mismo tiempo, lo que puede llevar a sesgo en las predicciones genómicas.

Con respecto a la primera restricción planteada en el punto 1, en nuestro primer trabajo, los estudios indicaron que el RFI fue genéticamente independiente de sus regresores, además de serlo fenotípicamente por definición. No obstante, tuvimos la inquietud de evaluar la equivalencia planteada en el punto 2 por Kennedy et al. (1993), al mismo tiempo de cuestionarnos y querer evaluar los puntos planteados en las restricciones 3 y 4.

Adicionalmente, nos pareció interesante investigar sobre modelos equivalentes que incluyan características predictoras con el fin de evaluar si estos modelos presentaban mejoras en cuanto a las precisiones individuales (analizado en el segundo artículo), así como mayor habilidad predictiva de los modelos (analizado en el tercer artículo).

Por lo tanto, el primer modelo alternativo al RFI clásico planteado para estudio fue una propuesta de estimación de un valor de cría para RFI obtenido directamente a partir de las características registradas, consumo (DMI), peso metabólico (MWT) y ganancia diaria (ADG) y espesor de grasa dorsal (BFAT), a través un modelo multivariado, donde se obtiene el DEP o valor de cría para cada una de ellas y posteriormente estos se relacionan en una combinación lineal (índice) para calcular un RFI *genético* (Crews et al., 2010; MacNeil et al., 2013). Esta alternativa fue demostrada por otros autores (Kennedy et al., 1993) como *equivalente*. La segunda alternativa estudiada fue la inclusión de peso al destete (PD) como característica indicadora de consumo y cuya información está disponible para un mayor número de animales. Esta alternativa fue analizada, ya que *a priori* presentaría ventajas por permitir eliminar el posible sesgo causado por la selección de los animales que participan de las pruebas de consumo (Pollak et al., 1984; Pszczola et al., 2013; Lu et al., 2015) y al mismo tiempo permitiría predecir el valor de cría a un mayor número de animales a través del uso de modelos multivariados. Para plantear los dos modelos

alternativos, primero se estimaron las correlaciones genéticas entre el DMI y sus regresores (MWT, ADG, BFAT) a través de un modelo multivariado, las cuales son presentadas en el segundo artículo. Para el tercer modelo, se estimaron las correlaciones genéticas entre DMI, regresores, y la característica predictora considerada, que fue el PD. A partir de dichas correlaciones, se obtuvieron los coeficientes de ponderación del índice a partir de las ecuaciones 4 y 5.

Los resultados de estos análisis, presentados y discutidos en el capítulo 4, que corresponde a un resumen expandido publicado en el WCGLAP, 2022 (Pravia et al., 2022b), confirman la cuarta hipótesis planteada en esta tesis. Estos análisis permitieron corroborar que la predicción de los valores de cría o DEP obtenidos para RFI con el modelo univariado y los obtenidos por el multivariado con las características regresoras fueron equivalentes, ya que no hubo reordenamiento de animales. Las correlaciones entre los DEP obtenidos por el modelo univariado y el modelo a partir de las propias variables regresoras que definen el RFI (ADG, MWT y BFAT) fueron de 0,99. En cuanto al tercer modelo, que considera el PD como característica predictora, las correlaciones de los DEP también fueron altas (0,96), aunque se observó una mayor dispersión de los DEP de este último modelo. Una de las principales ventajas esperables del uso de los modelos multivariados respecto a los univariados es la obtención de mejoras en las precisiones individuales (Mrode, 2005). Para ello, se compararon las precisiones individuales (*accuracy*) de los tres modelos por grupo de individuos. Se pudo constatar que los modelos multivariados presentaron precisiones más altas para el grupo de animales con fenotipo y especialmente para animales sin fenotipo, padre de datos o animales contemporáneos sin fenotipo que de otra manera no alcanzaban niveles de precisión mínimos. El incremento en las precisiones fue entre un 2,5 y un 6 % para el modelo multivariado equivalente respecto al univariado y de un 40 % aproximadamente para el modelo multivariado que incluye el PD.

El mayor incremento observado en las precisiones individuales del modelo que incluye el PD es consecuencia de la inclusión en el multivariado de una característica de alta heredabilidad como lo es el peso al destete ($h^2 = 0,35$, estimada en este trabajo) y genéticamente correlacionada con el consumo, ganancia, peso y espesor de

grasa. Las correlaciones genéticas entre el peso al destete y las características proxy fueron PD-DMI, 0,54; PD-ADG, 0,57; PD-MWT, 0,66 y PD-BFAT de 0,40 (Pravia et al., 2022a). Estos resultados son coincidentes con lo reportado en la bibliografía, en donde el uso de características correlacionadas en los programas de mejoramiento genético ha sido una de las alternativas mayormente utilizadas para seleccionar en características de baja heredabilidad o de difícil medición (Aguilar et al., 2011; Calus y Veerkamp, 2011; Guo et al., 2014; Manzanilla-Pech et al., 2017) o cuando la población de referencia es limitada (Pszczola et al., 2013).

Los resultados de este trabajo indican que en los DEP obtenidos por los tres modelos son similares, sin cambios de ranking entre los animales, debido a las altas correlaciones entre las estimaciones de los diferentes modelos. Adicionalmente, se observa una ventaja en términos de mayores precisiones individuales para los DEP generados a partir de los modelos multivariados. A partir de estos resultados, el siguiente cuestionamiento fue si esta información era suficiente para concluir que los modelos multivariados se comportan mejor en predecir los valores de cría para RFI en situaciones en donde la información genómica está disponible. Por esta razón, se estudió la habilidad predictiva de los modelos a través de validaciones cruzadas para evaluar la capacidad predictiva de estos en un contexto en donde la información genómica tiene un rol determinante.

6.3. Estudio de la habilidad predictiva de los modelos genómicos propuestos

Las evaluaciones genéticas tienen el principal objetivo de predecir de forma precisa los valores de cría o DEP para todos los candidatos a la selección. Para ello la metodología mayormente utilizada son las ecuaciones de modelos mixtos, a través de metodologías BLUP, las cuales utilizan distinto tipo de información, fenotipos, relaciones de pedigrí y genotipos (Aguilar et al., 2010, Aguilar et al., 2018). En el caso de características de difícil medición, tradicionalmente se han empleado características indicadoras o registros en progenies (Calus y Veerkamp, 2011). En este tipo de características, la inclusión de la información genómica ha generado un avance muy importante, ya que ha permitido predecir valores de cría con buenas precisiones a edades tempranas a pesar de no contar con el fenotipo. Sin embargo, ha

puesto de manifiesto algunos cuestionamientos sobre sesgo en las predicciones, lo que puede llevar a respuestas a la selección no deseadas (Macedo et al., 2020). En este sentido, el estudio de la capacidad predictiva pasa a tener un rol muy importante.

En nuestro trabajo se pudo comprobar que la incorporación de la información genómica a las evaluaciones genéticas permitió arrojar valores de cría para aquellos animales sin fenotipo e incrementó las precisiones individuales tanto para animales que contaban con el fenotipo como para padres o abuelos de los datos (Pravia et al., 2021, presentado en anexo). Este incremento en las precisiones individuales o *accuracy* se refiere a la calidad con que ese valor de cría fue predicho. Sin embargo, desde el punto de vista poblacional, el mayor interés está en evaluar la capacidad predictiva de los modelos. Esto se estudia a través de validaciones cruzadas, en donde a un set de datos se le borra el fenotipo y se lo predice con otra información. En caso de predicciones genómicas, se evalúa la capacidad predictiva haciendo uso únicamente de la información genómica.

El tercer objetivo de la tesis, que engloba los anteriores, fue evaluar la capacidad predictiva de dos modelos genómicos propuestos para predecir valores de cría para RFI en toros Hereford. El primer modelo evaluado fue el univariado, actualmente utilizado. El segundo fue el modelo RFI propuesto y discutido previamente, el cual fue obtenido a partir del modelo multivariado que combina los EBV individuales para las características regresoras medidas en la prueba: consumo de alimento, peso metabólico, ganancia diaria y espesor de grasa dorsal, donde además se incluyó el peso al destete como característica predictora.

Para evaluar la capacidad predictiva de los modelos, se consideraron los indicadores propuestos por Legarra y Reverter (2018), adaptados por Bermann et al. (2021). Dichos indicadores se basan en la comparación de sucesivas evaluaciones genéticas *nuevas* o *parciales* y las *completas* o *totales*, lo que permite obtener varios estimadores del sesgo, dispersión y precisiones de las predicciones. Esta teoría se fundamenta en que las predicciones de las nuevas generaciones sin fenotipo podrían estar sesgadas como consecuencia de la selección (Macedo et al., 2020). En nuestra base de datos de 780 individuos Hereford, no contábamos con suficientes generaciones para evaluar sucesivas particiones, con lo cual planteamos algunas

alternativas para particionar los datos y obtener mayor número de réplicas. De esta manera, se consideraron dos estrategias de partición de los datos para realizar las validaciones, partición por réplicas al azar y partición por pruebas. La primera consistió en realizar treinta réplicas independientes al azar, en donde en cada réplica se eliminó el fenotipo a $\frac{1}{4}$ de los datos (set de validación o focal) y se predijo el mérito genético para ese grupo de individuos a partir de su genotipo y la población de referencia formada por $\frac{3}{4}$ de la población restante (set de entrenamiento). Esto se repitió treinta veces y para cada réplica se calcularon los indicadores propuestos: el sesgo ($B0$), la dispersión ($B1$), relaciones de precisión ($\rho_{w,p}$) y el incremento relativo por el agregado de fenotipos (IP), los que posteriormente se promediaron para el total de réplicas. En la segunda estrategia, se borraban los fenotipos para una de las nueve pruebas por vez y se predecían los valores de cría de los individuos de cada prueba a partir de su genotipo y la población de entrenamiento formada por todas las pruebas menos la prueba por validar. En cada ronda se obtuvieron los indicadores de habilidad predictiva.

Los resultados fueron publicados en Pravia et al. (2023) y mostraron que ambos modelos genómicos tuvieron una buena capacidad predictiva cuando los fenotipos eran removidos. Esto comprueba la tercera hipótesis en cuanto a que el uso de la información genómica en la predicción del mérito genético para RFI de los animales sin fenotipo es una herramienta que permite predecir el valor de cría para dichos animales de forma precisa. Los indicadores arrojaron que las predicciones fueron insesgadas ($B0$ entre -0,018 y 0,020 para el modelo univariado y entre -0,19 y 0,037 para el multivariado). En lo que respecta al indicador que evalúa la dispersión ($B1$), el modelo univariado se comportó mejor que el multivariado (0,96 vs. 0,89 en las validaciones de las treinta réplicas o en las validaciones por prueba). Estos valores menores a 1 indican una leve sobreestimación de las predicciones, ya que el valor esperado es 1. En lo que respecta al indicador *ratio of accuracy*, $\rho_{w,p}$, que mide la correlación entre la predicción para el set de individuos con la base completa ($\hat{u}_w$), en comparación con la predicción basada solo en el genotipo ($\hat{u}_p$), el valor promedio para el modelo univariado fue de 0,59 y 0,62 para la estrategia de validación al azar y por pruebas, respectivamente, mientras que, para el modelo multivariado, el valor

promedio fue de 0,51 para las validaciones al azar y 0,52 para las validaciones por prueba, respectivamente. En el caso del incremento en precisiones por el agregado de fenotipos (*IP*), que se deriva del indicador anterior, se obtuvo un valor promedio de 71 % para el modelo univariado, mientras que para el multivariado fue de 96 %. La interpretación de este indicador sugiere que, a medida que incorporan más fenotipos, se observa un incremento significativo en las precisiones, especialmente para el modelo multivariado. Considerando el conjunto de indicadores obtenidos, se concluye que las predicciones genómicas para RFI fueron más precisas en el modelo univariado que en el modelo multivariado propuesto.

Se pudo constatar que la conformación de la población parcial o focal y la población de referencia o total influyó en los resultados de habilidad predictiva. Cuando las validaciones fueron realizadas con el set elegido al azar, donde cada réplica de set de datos parcial contiene animales de distintas pruebas, se comportó mejor, mientras que, cuando las validaciones se realizaron por prueba, los indicadores de habilidad predictiva difirieron bastante según la prueba. En algunas pruebas (1, 7 y 8), los indicadores demostraron una buena capacidad predictiva, mientras que en otras, como por ejemplo la 5, los indicadores fueron más bajos. Este comportamiento fue similar tanto para el modelo univariado como para el multivariado, aunque en todos los casos los indicadores de habilidad predictiva fueron inferiores para el modelo multivariado que el univariado. La variabilidad entre pruebas es coincidente con lo reportado en la bibliografía en cuanto a que las predicciones son sensibles a la conformación del grupo de validación y las relaciones de parentesco entre los animales que pertenecen al grupo de entrenamiento y al de validaciones (Saatchi et al., 2011; Pszczola et al., 2012). Estos resultados llevaron a cuestionar que los animales de algunas pruebas estuvieran genéticamente más distanciados y esto pudiera haber repercutido en peores predicciones. Sin embargo, al realizar un estudio de componentes principales, no se observó ninguna tendencia de agrupamiento entre los animales de diferentes pruebas.

Contar con animales más homogéneos permite tener predicciones más precisas. Por otro lado, mantener una población de entrenamiento dinámica con incorporación de nuevos genotipos permitiría mantener un parentesco genético constante o

adecuado entre los animales que conforman la población de entrenamiento y los futuros candidatos a la selección (Saatchi et al., 2011).

La validación a través de las pruebas representa la situación más cercana a la realidad en la medida que cada prueba pueda representar el ingreso de una nueva generación de individuos a predecir. En el modelo univariado, las predicciones genómicas para los candidatos sin fenotipo se basan en la información provista por la población de referencia, la que está conformada por los fenotipos de individuos de esa generación más individuos de generaciones anteriores. En el modelo multivariado, las predicciones se basarían en los fenotipos de los contemporáneos y anteriores medidos en eficiencia, así como en los pesos al destete de todos los individuos contemporáneos.

Nuestra quinta hipótesis planteaba que los modelos multivariados mostrarían un mejor desempeño al hacer uso de la información de una característica correlacionada. Sin embargo, los resultados no respaldaron esta suposición. Investigadores reportaron mejores predicciones utilizando características indicadoras o proxy en modelos multivariados genómicos (Calus y Veerkamp, 2011; Guo et al., 2014; Jibila et al., 2021; Pszczola et al., 2012). Sin embargo, algunos investigadores (Manzanilla-Pech et al., 2017) observaron que, cuando la característica a predecir está presente en multivariado, no se observan beneficios por incluir las características predictoras. A su vez, Calus y Veerkamp (2011) indicaron que los resultados son dependientes de las magnitudes de las correlaciones genéticas entre las características predictoras y la característica por predecir. Si la correlación es alta, mayor a 0,5, el uso de características indicadoras se refleja en una mejor predicción, mientras que, si la correlación es baja, se obtienen mayores beneficios por genotipar el individuo que por obtener el fenotipo de la característica correlacionada. En los trabajos mencionados, la característica por predecir era consumo de alimento. Sin embargo, en nuestro estudio, la característica de interés es RFI, lo que introduce una complejidad adicional, ya que no se trata de un rasgo directamente medido, sino del residuo de una regresión lineal. Nuestra premisa fue que, dado que el peso al destete está altamente correlacionado con dichas características, su inclusión en un modelo multivariado podría mejorar la predicción de los DEP individuales de estas

características. Consecuentemente, la obtención del RFI como un índice de selección a partir de DEP más precisos resultaría en una mejor predicción de este indicador. Sin embargo, en ausencia de fenotipo de consumo y características de la prueba, el incluir los fenotipos de PD en el modelo como únicos predictores no generó un beneficio en la habilidad predictiva cuando los genotipos estaban presentes. En situaciones en donde no se contara con genotipos y las predicciones estuvieran basadas únicamente en fenotipos, los resultados podrían ser diferentes. Sin embargo, este escenario no fue objeto de nuestro estudio, ya que es improbable debido a que la información genómica está siendo cada vez más utilizada para predecir los valores de cría en este tipo de características de difícil medición. A esto se suma la promoción del uso de información genómica como forma de acelerar la tasa de progreso genético en todas las características de importancia incluidas en las evaluaciones genéticas.

7. Conclusiones

El primer resultado de este trabajo demostró que tanto el consumo de alimento como el RFI presentan variabilidad genética y heredabilidades de medias a altas para las poblaciones estudiadas. Esto indica que es posible obtener respuestas a la selección por estas características. Con respecto al consumo, los parámetros genéticos confirman su alta correlación positiva con el peso, ganancia diaria y espesor de grasa, por lo que la selección por menor consumo traerá aparejadas respuestas no deseadas en estas características. En el caso de RFI, las correlaciones genéticas estimadas muestran que el RFI es genéticamente independiente de esas características. Estos resultados indican una ventaja en el uso del RFI sobre el consumo como criterio de selección, ya que la selección y apareamiento de individuos a través de menores RFI se verá reflejado en progenies con menor consumo sin afectar negativamente el desempeño productivo en otras características de importancia económica.

Considerando el RFI como la característica definida en los programas de mejora genética nacionales para el incremento de la eficiencia, se propusieron y evaluaron otros modelos genéticos multivariados alternativos, bajo la hipótesis de que tendrían algunas ventajas. Con esta propuesta se obtuvieron incrementos en las precisiones individuales de los valores de cría para el RFI. Sin embargo, estas ventajas no se reflejaron en mejores habilidades predictivas del modelo multivariado a nivel poblacional, ya que los indicadores analizados demostraron que el modelo univariado tuvo mayor capacidad predictiva cuando no se contó con información fenotípica en los candidatos.

Si bien los resultados obtenidos en este sentido no corroboraron la quinta hipótesis planteada, este trabajo realizó una importante contribución al desarrollo metodológico e implementación del uso de modelos multivariados para eficiencia de conversión, los cuales están siendo considerados a nivel internacional. Los antecedentes establecidos en este estudio pueden ser retomados e incorporados en propuestas futuras para la construcción de índices de selección que incluyan el consumo de alimento, así como RFI.

En un contexto global en el que la inclusión de la información genómica en las evaluaciones genéticas está creciendo rápidamente, evaluar la capacidad predictiva de los modelos que se están utilizando a nivel nacional cumple un rol fundamental para ampliar el uso de esta herramienta. En este sentido, el segundo resultado más significativo de este estudio fue la demostración de que es posible predecir, de forma precisa, valores de cría para la característica RFI en animales candidatos a la selección a partir de sus genotipos. Esto tiene gran relevancia a nivel nacional, ya que permite expandir el uso de esta herramienta a aquellos candidatos a la selección, a los cuales no es posible medirles su consumo de alimento ni, por ende, estimar el RFI.

Los resultados de este trabajo refuerzan la importancia de seguir fortaleciendo la población de entrenamiento a los efectos de asegurar buenas conexiones genómicas entre los individuos que la integran y los candidatos a la selección. Se comprobó cómo la habilidad predictiva de las predicciones fue variable según los grupos de individuos por predecir. Esto reafirma la necesidad de continuar midiendo y genotipando individuos contemporáneos y vinculados con los candidatos a predecir.

8. Implicancias

El consumo de alimento y la eficiencia de conversión han sido reconocidos como caracteres de alta relevancia económica en los sistemas ganaderos por su impacto directo sobre los costos de los sistemas de producción, así como por la relación directa que tiene sobre las emisiones de gases de efecto invernadero.

La selección genética es una herramienta efectiva para la mejora de la eficiencia de conversión y sostenibilidad de la ganadería. La principal limitante para este tipo de características ha sido disponer de equipos para medir consumo y medir a todos los candidatos a la selección. Gracias a la implementación de selección genómica y la construcción de poblaciones de entrenamientos para este tipo de características, es que hoy es posible predecir DEPs para consumo y RFI con buenas precisiones al total de candidatos a la selección, lo cual ha sido demostrado en este trabajo.

Contar con esta herramienta de selección posiciona a la ganadería de Uruguay, a un nivel de vanguardia, por su alto impacto a nivel productivo, económico y ambiental. La implementación de la selección genómica en estas características a mayor escala permitirá disponer de una ganadería más eficiente en el uso de recursos, contar con animales con iguales niveles de producción de carne, y con un menor consumo de alimento y menores emisiones de metano.

Los próximos desafíos desde la investigación es seguir apostando a incrementar las poblaciones de entrenamiento para estas características, de manera de poder generar predicciones genómicas cada vez más precisas. Con un mayor volumen de datos se podrían estimar las correlaciones genéticas con otras características de interés como peso a los 18 meses, características carniceras, que en este trabajo no fue atendido por no contar con la información suficiente al momento de la tesis. También continuar fortaleciendo los estudios de relacionamiento entre consumo de alimento y emisiones de metano, los cuales en nuestro país se han

comenzado hace algunos años con resultados alentadores, pero requiere seguir profundizando.

Otro desafío que tienen los técnicos extensionistas es trabajar en la difusión y en mayor adopción de estos nuevos DEPs por parte de los criadores usuarios, así como promover programas de certificación o valorización de productos, a partir de la utilización de una ganadería que produce animales con igual desempeño productivo, en cortes de valor y aptitudes de calidad de carne, consumiendo menos kg de alimento y generando una menor huella ambiental por producir menores emisiones de metano.

9. Bibliografía

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S. y 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.
- Aguilar, I., Misztal, I., Legarra, A. y Tsuruta, S. (2011). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *Journal of Animal Breeding and Genetics*, 128(6), 422-428.
- Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A. y Misztal, I. (2018). BLUPF90 suite of programs for animal breeding with focus on genomics. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, 11.751.
- Archer, J. A., Barwick, S. A., Grase y H. -U. (2004). Economic evaluation of beef cattle breeding schemes incorporating performance testing of young bulls for feed intake. *Australian Journal of Experimental Agriculture*, 44 (5):393-404.
- Arthur, P. F., Archer, J. A., Johnston, D. J., Herd, R. M., Richardson, E. C. y Parnell, P. F. (2001). Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *Journal of Animal Science*, 79, 2805-2811.
- Basarab J. A., Beauchemin K. A., Baron V. S., Ominski K. H., Guan L. L., Miller S. P. y Crowley J. J. (2013). Reducing GHG emissions through genetic improvement for feed efficiency: effects on economically important traits and enteric methane production. *Animal*, 7 (suppl. 2): 303-315. 10.1017/S1751731113000888.
- Basarab J. A., Price M. A., Aalhus J. L., Okine, E. K., Snelling, W. M., Lyle y K. L. (2003). Residual feed intake and body composition in young growing cattle. *Canadian Journal of Animal Science*, 83, 189-204.
- Beef Improvement Federation. (2010). *Guidelines for Uniform Beef Improvement Programs* (9.^aed.). https://beefimprovement.org/wp-content/uploads/2013/07/BIFGuidelinesFinal_updated0916.pdf

- Bermann, M, Legarra, A, Hollifield, M. K., Masuda, Y., Lourenco, D. y Misztal, I. (2021). Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. *Journal of Animal Breeding and Genetics*, 138(1), 4-13. 10.1111/jbg.12507.
- Berry, D. P. y Crowley, J. J. (2013). Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. *Journal of Animal Science*, 91(4), 1594-613.
- Berry, D. P., Coffey, M. P., Pryce, J. E., de Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J. J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K. y Veerkamp, R. F. (2014). International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *Journal of Dairy Science*, 97(6), 3894-905. doi:10.3168/jds.2013-7548.
- Calus M. P. L. (2010). Genomic breeding value prediction: methods and procedures. *Animal*, 4(2), 157-164.
- Calus, M. P. L. y Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetic Selection Evolution*, 43(1), 26. doi: 10.1186/1297-9686-43-26.
- Calus, M. P. L., Berry, D. P., Banos, G., de Haas, Y. y Veerkamp, R. F. (2013). Genomic selection: the option for new robustness traits? *Advances in Animal Biosciences*, 4(3), 618-625.
- Calus, M. P. L., de Haas, Y., Pszczola, M. y Veerkamp, R. F. (2013). Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Animal*, 7, 183-191.
- Chen, L., Schenkel, F., Vinsky, M., Crews, D. H. y Li, C. (2013). Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. *Journal of Animal Science*, 91, 4669-4678.
- Crews, D. H. Jr. (2005). Genetic of efficient feed utilization and national cattle evaluation: a review. *Genetics and Molecular Research*, 4, 152-165.
- Crews, D. H., Pendely, C. T., Carstens, G. E. y Mendes, E. D. M. (2010). Genetic characterization of feed intake and utilization in performance tested beef bulls. *Proceeding Western Section, American Society of Animal Production*. Vol. 61.

- Gianola, D., de los Campos, G., Hill, W.G., Manfredi, E., Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics*. Sep;183(1):347-63. 10.1534/genetics.109.103952.
- Habier, D.; Fernando, R.L., Dekkers, J.C.M. (2007). The impact of genetic relationship information of genome-assisted breeding values. *Genetics* 177:2389-2397.
- Habier, D., Tetens, J., Seefried, F. R., Lichtner, P. y Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, 42(5). 10.1186/1297-9686-42-5.
- Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., Keele, J. W., Smith, T. P., Chitko-McKown, C .G. y Laegreid, W. W. (2002). Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome*, 13(5), 272-281. 10.1007/s00335-001-2146-3.
- Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, 31(2), 423-447. <https://doi.org/10.2307/2529430>
- Herd, R. M., Archer, J. A. y Arthur, P. F. (2003). Reducing the cost of beef production through genetic improvement in residual feed intake: Opportunity and challenges to application. *Journal of Animal Science*, 81(13), 9-17.
- Herd, R. M. y Bishop, S.C. (2000). Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livestock Production Science*, 63, 111-119.
- Kennedy, B. W., J. H. J. Van der Welf y T. H. E. Meuwissen. (1993). Genetic and statistical properties of residual feed intake. *Journal of Animal Science*, 71, 3239-3250.
- Koch, R. M., Swiger, L. A., Chambes, D. y Gregory, K. E. (1963). Efficiency of feed use in beef cattle. *Journal of Animal Science*, 22, 486-494.
- Legarra, A. (2018). Course Notes Risks of forecasting in Animal Breeding. https://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=andres_part6.pdf

- Legarra, A., Aguilar, I. y Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9), 4656-4663. 10.3168/jds.2009-2061.
- Legarra, A., Christensen, O. F., Aguilar, I. y Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science*, 166, 54-65.
- Legarra, A. y Reverter, A. (2017). Can we frame and understand cross-validation results in animal breeding? *Proc. Assoc. Advmt. Anim. Breed. Genet.*, 22, 73-80
- Legarra, A. y Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetic Selection Evolution*, 50, 53.
- Legarra, A., Robert-Granie, C. R., Manfredi, E. y Elsen, J.M. Performance of Genomic Selection in Mice. (2008). *Genetics*, 180(1), 611-619.
- Lu, Y., Akanno, D. E. C., Crowley, J. J., Schenkel, F., Li, H., De Pauw, M., Moore, S. S., Wang, Z., Li, C., Stothard, P., Plastow, G., Miller, S. P. y Basarab, J. A. (2016). Accuracy of genomic prediction for feed efficiency traits of beef cattle using 50K and imputed HD genotypes. *Journal of Animal Science*, 94, 1342-1353.
- Lu, Y., Vandehaar, M. J., Spurlock, D. M., Weigel, K. A., Armentano, L. E., Staples, C. R., Connor, E. E., Wang, Z., Bello, N. M. y Tempelman, R. J. (2015). An alternative approach to modeling genetic merit of feed efficiency in dairy cattle. *Journal of Dairy Science*, 98(9), 6535-6551. 10.3168/jds.2015-9414.
- Macedo, F. (2021). *Study of the origin and impact of bias in national genetic and genomic evaluations. The case of dairy sheep* [tesis doctoral]. Institut National Polytechnique de Toulouse.
- Macedo, F., Reverter, A. y Legarra. (2020). Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *Journal of Dairy Science*, 103(1), 529-544.
- MacNeil, M. D., Lopez-Villalobos, N. y Northcutt, S. L. (2011). A prototype national cattle evaluation for feed intake and efficiency of Angus cattle. *Journal of Animal Science*, 89, 3917-3923.

- MacNeil, M. D., Scholz, M. M. y Maiwashe, A. (2013). Estimates of variance components for postweaning feed intake and growth in Bonsmara bulla and evaluation of alternatives measures of feed efficiency. *South African Journal of Animal Science*, 43, 18-24.
- Manzanilla-Pech, C. I. V., Veerkamp, R. F., de Haas, Y., Calus, M. P. L. y Napel, J. ten. (2017). Accuracies of breeding values for dry matter intake using non genotypes animals and predictor traits in different lactations. *Journal of Dairy Science*, 100, 9103-9114.
- Mantysaari, E. A., Liu, Z. y VanRaden, P. (2010). Interbull validation test for genomic evaluations. *Interbull Bull*, 41, 17-21.
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetic Selection Evolution*, 41, 35
- Meuwissen, T., y Goddard, M. (1996). The use of marker haplotypes in animal breeding schemes. *Genetic Selection Evolution*, 28, 161-176.
- Meuwissen, T. y Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, 185(2), 623-31. 10.1534/genetics.110.116590.
- Meuwissen, T. H. E., Hayes, B. J. y Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.
- Misztal, I., Legarra, A. y Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92(9), 4648-4655. 10.3168/jds.2009-2064.
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T. y Lee, D.H. (2002). BLUPF90 and related programs (BGF90). *Proceedings of the 7th world congress on genetics applied to livestock production* (pp. 19-23), Montpellier.
- Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A. y Vitezica, Z. (2016). Manual for BLUPF90 family of programs. https://nce.ads.uga.edu/html/projects/programs/docs/blupf90_all8.pdf
- Mrode, R. A. (2005). *Linear Models for the Prediction of Animal Breeding values* (2.^a ed.). CABI Publishing.

- Navajas, E., Ravagnolo, O., De Barbieri, I., Pravia, M. I., Aguilar, I., Lema, O. M., Vera, B., Peraza, P., Marques, C. B., Velazco, C. B. y Ciappesoni, G. (2022). *Proceedings of the World Congress on Genetics Applied to Livestock Production (WCGALP), 12*, Rotterdam.
- Nejati-Javaremi, A., Smith, C. y Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75(7), 1738-1745. 10.2527/1997.7571738x.
- Nielsen, M. K., MacNeil, M. D., Dekkers, J. C. M., Crews, D. H. Jr., Rathje, T. A., Enns, R. M. y Weaber, R. L. (2013). Review: Life-cycle, total industry genetic improvement of feed efficiency in beef cattle: Blueprint for the Beed Improvement Federation. *The Professional Animal Scientist*, 29, 559-565.
- Pollak, E. J., Van der Werf, J y Quaas, R. L. (1984). Selection bias and multiple trait evaluation. *Journal of Dairy Science*, 67, 1590-1595.
- Pszczola, M., Strabel, T., Van Arendonk, J. A. M. y Calus, M. P. L. (2012). The impact of genotyping different group of animals on accuracy when moving from traditional to genomic selection. *Journal of Dairy Science*, 95, 5412-5421.
- Pszczola, M., Veerkamp, R. F., de Hass, Y., Wall, E., Strabel, T. y Calus, M. P. L. (2013). Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal*, 7, 11.
- Pravia, M. I., Ravagnolo, O., Urioste, J. y Garrick, D. J. (2014). Identification of breeding objectives using a bioeconomic model for a beef cattle production system in Uruguay. *Livestock Science*, 160, 21-28. <https://doi.org/10.1016/j.livsci.2013.12.006>
- Pravia, M. I., Ciappesoni, G., Aguilar, I., Ravagnolo, O., Lema, M. y Navajas, E. A. (2015). Analysis of feed intake and feed efficiency in Hereford Bulls and steers. *52a Reunião Anual da Sociedade Brasileira de Zootecnia. Belo Horizonte*.
- Pravia, M. I., Navajas, E. A., Aguilar, I. y Ravagnolo, O. (2021). Incorporación de información genómica en la estimación del mérito genético de consumo

- residual de alimento. *Archivos Latinoamericanos de Producción Animal*, 29 (supl. 1), 74-76.
- Pravia, M. I., Navajas, E., Aguilar, I. y Ravagnolo, O. (2022a). Evaluation of feed efficiency traits in different Hereford populations and their effect on variance component estimation. *Animal Production Science*, 62(17), 1652-1660. <https://doi.org/10.1071/AN21420>
- Pravia, M. I., Navajas, E., Aguilar, I. y Ravagnolo, O. (2022b). Alternative models to predict residual feed intake in Hereford breed and effects on their breeding values accuracy. Part 6 - Challenges - resource allocation and genetics of feed intake and efficiency. *Proceedings of the World Congress on Genetics Applied to Livestock Production (WCGALP)*, 12, Rotterdam. <https://inia.uy/sites/default/files/publications/2024-10/978-90-8686-940-4-48.pdf>
- Pravia, M. I., Navajas, E., Aguilar, I. y Ravagnolo, O. (2023). Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake. *Journal of Animal Breeding and Genetics*, 140(5), 508-518. <https://doi.org/10.1111/jbg.12775>
- Ravagnolo, O., Aguilar, I., Crowley, J. J., Pravia, M. I., Lema, M., Macedo, F. L., Scott, S. y Navajas, E. A. (2018). Accuracy of genomic predictions of residual feed intake in Hereford with Uruguayan and Canadian training populations. *Proceedings of the World Congress on Genetics Applied to Livestock Production (WCGALP)*, Volume Electronic Poster Session – Species – Bovine (beef); Auckland, New Zealand. 1, p. 723.
- Saatchi, M., Miraei-Ashtiani, S. R., Nejati Javaremi, A., Moradi-Shahrehabak, M. y Mehrabani-Yeghaneh, H. (2010). The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *African Journal of Biotechnology* Vol. 9 (4), pp. 438-442.
- Saatchi, M., McClure, M. C., McKay, S. D., Rolf, M. M., Kim, J., Decker, J. E., Taxis T. M., Chapple, R. H., Ramey, H. R., Northcutt, S. L., Bauck, S. (2011).

- Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution*, 43, 40.
- Silva, R. M., Fragomeni, B. O., Lourenco, D. A., Magalhães, A. F., Irano, N., Carneiro, R., Canesin, R. C., Mercadante, M. E., Boligon, A. A., Baldi, F. S., Misztal, I. y Albuquerque, L. G. (2016). Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. *Journal of Animal Science*, 94(9), 3613-3623. 10.2527/jas.2016-0401.
- Sargolzaei, M., Chesnais, J. P. y Schenkel, F. S. (2011) Fimpute - an efficient imputation algorithm for dairy cattle populations. *Journal of Dairy Science*, 94 Suppl 1421 (333).
- Tempelman, R. J., Spurlock, D. M., Coffey, M., Veerkamp, R. F., Armentano, L. E., Weigel, K. A., de Hass, Y., Staples, C. R., Connor, E. E., Lu, Y. y VanderHaar, M. J. (2015). Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *Journal of Dairy Science*, 98, 2013-2026.
- Tsuruta, S., Misztal, I., Aguilar, I. y Lawlor, T. J. (2011). Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US. Holsteins. *Journal of Dairy Science*, 94, 4198-4204.
- VanderWerf, J. H. J. (2004). It is useful to define residual feed intake as a trait in animal breeding programs? *Australian Journal of Experimental Agriculture*, 44, 405-409.
- VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414-4423.
- VanRaden, P. M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. y Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24.
- Villanueva, B., Pong-Wong, R., Fernandez, J. y Toro, M. A. (2005). Benefits from marker assisted selection under an additive polygenic genetic model. *Journal of Animal Science*, 83(8), 1747-1752.

Vitezica, Z. G., Aguilar, I., Misztal, I. y Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genetic Research*, 93, 357-366.

10. Anexos

10.1 Incorporación de información genómica en la estimación del mérito genético de consumo residual de alimento.



Congreso de la Asociación Uruguaya de Producción Animal
14-15 de Diciembre de 2021
Congreso virtual

Incorporación de información genómica en la estimación del mérito genético de consumo residual de alimento

Pravia, M.I.^{1*}, Navajas, E.A.¹, Aguilar, I.¹, Ravagnolo, O.¹

¹Instituto Nacional de Investigación Agropecuaria, Las Brujas, Rincón del Colorado, Canelones (Uruguay).
*mpravia@inia.org.uy

Introducción

El consumo residual de alimento (RFI, *residual feed intake*), es la diferencia entre el consumo actual y el consumo predicho basado en requerimientos de mantenimiento y producción (Koch et al., 1963). Su incorporación en los programas de mejora ha sido lenta por su difícil medición y alto costo. En consecuencia, el número de animales evaluados para esta característica es bajo. La incorporación de información genómica permitiría mejorar las precisiones de los valores de cría, y contando con una población de referencia conectada, obtener predicciones genómicas precisas para candidatos a la selección que no cuenten con el fenotipo (Ravagnolo et al., 2018). El objetivo de este trabajo es el estudio del efecto de la incorporación de información genómica sobre las predicciones del mérito genético para RFI.

Materiales y métodos

La base de datos consta de registros de 780 toros Hereford evaluados en pruebas de eficiencia en el postdestete, recolectados a partir del proyecto RTS_1_2012_1_3489 financiado por ANII. Estas se basaron en protocolos de registración internacional y consistieron en 70 días de evaluación donde se registró el consumo de alimento individual con un sistema automatizado (Growsafe). En cada prueba se obtuvo información individual de consumo de alimento diario promedio (CMD, kgMS/día), ganancia diaria durante la prueba calculada por regresión (GD, kg/día), peso metabólico como peso a mitad de prueba^{0.75} (PMT), espesor de grasa dorsal medido por ultrasonografía (EGS, mm), para el cálculo del RFI como:

$$CMD = CG + b_1 GD + b_2 PMT + b_3 EGS + e$$

donde CG es el grupo contemporáneo al cual pertenece el animal (año, prueba, corral), CMD, GD, PMT, EGS fueron definidos previamente, y b_1 , b_2 y b_3 los coeficientes de regresión de cada característica sobre el consumo. El residuo representa el RFI.

El modelo para la predicción de los méritos genéticos de RFI es:

$$Y = Xb + Zu + e \quad [1]$$

donde Y son las observaciones de RFI, b los efectos fijos del modelo, u es el vector del efecto genético del animal para RFI, X es la matriz que relaciona las observaciones con los efectos fijos (edad del animal a inicio de la prueba, y edad de la madre), y Z con los efectos aleatorios (efecto animal). Se asume que $u \sim N(0, A\sigma_{uRFI}^2)$, siendo A la matriz de parentesco aditivo entre los animales, σ_{uRFI}^2 la varianza genética aditiva de RFI (0.08 obtenida en Pravia et al., 2021) y $e \sim N(0, I\sigma_{eRFI}^2)$.

Para la predicción de los valores de cría genómicos se reemplaza la matriz A por la matriz de parentesco H, que combina el parentesco genealógico y el parentesco basado en información molecular (Aguilar et al., 2010), permitiendo incorporar información genómica en la predicción de los méritos genéticos. Las soluciones para las predicciones del mérito genético fueron obtenidas utilizando BLUP y GBLUP implementadas en BLUPF90 (Miszta et al., 2016). Las precisiones (acc) de las estimaciones fueron obtenidas a través de la varianza del error de la predicción (PEV) como $1 - PEV / (1 + F)\sigma_g^2$. Para ambos análisis, se consideraron los fenotipos mencionados, 19.009 animales extraídos del pedigree para el modelo aditivo, y 26.572 animales extraídos del pedigree y 5439 animales con genotipados imputados al panel de 50k SNP (BovineSNP50, Illumina San Diego, CA) para el modelo genómico.



Congreso de la Asociación Uruguaya de Producción Animal
14-15 de Diciembre de 2021
Congreso virtual

Resultados y discusión

Las predicciones de los valores de cría genómicos (Gebv_rfi) estuvieron altamente correlacionadas ($r=0,95$) con las predicciones considerando el parentesco aditivo (Ebv_rfi), tanto para aquellos animales que cuentan con fenotipo (Figura 1), así como para el total de la población ($r=0,90$). El reemplazar la matriz A por la matriz H aumentan las precisiones (Figura 2), pasando de 0,305 (Acc_rfi) a 0,334 (Gacc_rfi) en los animales con fenotipo. También se observaron incrementos en las precisiones de animales parientes de aquellos con dato, como padres (con o sin genotipo) y animales contemporáneos que no cuentan con fenotipo de RFI, pero sí con genotipo (Cuadro 1). Estas mejoras se deben a una mejor estimación del parentesco real entre los animales genotipados considerados en la matriz H (Aguilar et al., 2010), la que consecuentemente es trasladada a todos los demás animales vinculados en el pedigree por parentesco aditivo.

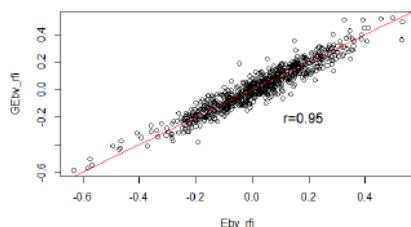


Fig. 1. Comparación de las predicciones de mérito genético considerando parentesco aditivo (Ebv_rfi) vs con la inclusión de genómica (Gebv_rfi).

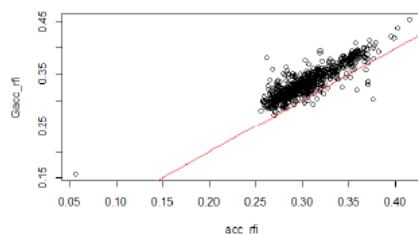


Fig. 2. Comparación de las precisiones de las predicciones con parentesco aditivo (acc_rfi) vs con la inclusión de genómica (Gacc_rfi).



Congreso de la Asociación Uruguaya de Producción Animal
14-15 de Diciembre de 2021
Congreso virtual

Cuadro 1. Impacto de la inclusión de información genómica sobre las predicciones del mérito genético (Ebv_rfi, vs Gbev_rfi) y sobre sus precisiones (Acc_rfi vs Gacc_rfi).

Grupo evaluado	Num	SIN GENÓMICA		CON GENÓMICA	
		Ebv_rfi	Acc_rfi	GEbv_rfi	Gacc_rfi
Animal con dato	778	0,018	0,305	0,014	0,334
Padre de dato 1 hijo y sin genotipo	113	0,009	0,139	0,008	0,151
Padre de dato con más de 1 hijo y con genotipo	87	0,003	0,215	0,001	0,262
Animal sin fenotipo y con genotipo	133	0,037	0,095	0,031	0,152

Conclusión

La incorporación de información genómica a través del reemplazo de la matriz de parentesco A por la matriz H, implica mejoras en las predicciones de los valores de cría para RFI. Esto es de especial importancia para aquellas características de difícil medición en las cuales se cuenta con escasa información fenotípica. Estos resultados serán complementados con estudios de la capacidad predictiva del modelo a través de validaciones.

Palabras Claves: RFI, GBLUP, precisiones.

Literatura citada

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. J. Dairy Sci. 93 :743–752
- Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2016. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf
- Pravia, M.I, Navajas, E., Aguilar, I., Ravagnolo, O. 2021. Enviado a Animal Production Science, Manuscript AN21420, en revision.
- Ravagnolo, O., Aguilar, I., Crowley, J.J., Pravia, M.I., Lema, M., Macedo, F.L., Scott, S., Navajas, E.A. 2018. Proc. WCGALP 11:723.
- Koch, R.M, L.A.Swiger, D.Chambes and K.E. Gregory. 1963. J. Anim. Sci. 22:486-494.