MÁSTER INTERUNIVERSITARIO EN MEJORA GENÉTICA ANIMAL Y BIOTECNOLOGÍA DE LA REPRODUCCIÓN

Analysis of feed efficiency in Holstein calves and adult cows using predictive models

Tesis de Máster

Valencia, Septiembre 2020

Edgar Leonardo Reinoso Pelaez

Director:

Dr. Oscar González Recio

Dedication

To my parents, who with example and affection have given me the education, dignified life and have shown me that true love exists. I will always admire their determination, perseverance and courage to face the challenges. Without them, nothing of this would be possible.

Edgar

Acknowledgments

This Master thesis was developed at the Animal Breeding department of the National Institute for Agricultural and Food Research and Technology (INIA) in Madrid, Spain.

I would like to thank my advisor Oscar González Recio, who with patience, charisma and impeccable leadership shared his knowledge throughout the research process. He has proven to be an outstanding professional and an extraordinary person.

I thanks to Alex Bach and IRTA for contributing the data and collaborating with this project.

I am grateful Agustin Blasco for the excellent organization of the master in Animal Breeding and Reproduction Biotechnology, and especially, for his talented and particular way of teaching. I also thank all the teachers who friendly shared their knowledge.

I am filled with gratitude with INIA and all the members of the Animal Breeding department for their great human quality. They have made me feel at home.

A special thanks to my parents Edgar Reinoso Rodriguez and Marianita Pelaez Zaruma. They covered the expenses of study and stay, and always they have been with me.

I thank my family and friends, who trusted me and constantly give me their support and affection.

To my friends in the Master degree, together we form a family and share moments that I will never forget.

Finally. Thank you, Carolina. I know that you are no longer in this world, but I want you to know that you were a great friend. We share several years of camaraderie in veterinary degree and heartfelt thanks for your support, trust and friendship. Wherever you are, I hope you have peace.

Table of Contents

Index of Tables

Index of Figures

RESUMEN

La demanda creciente de alimento debido al aumento de la población requiere de una ganadería más eficiente y sostenible. En la ganadería lechera el 50-70% de los costos de producción se deben a la alimentación, por lo que es necesario optimizar los procesos y seleccionar animales con alto nivel productivo y bajo consumo, es decir, con alta eficiencia alimentaria (EA). No todas las granjas pueden estimar EA en su rebaño porque su medición es costosa; por ello, la genética juega un papel importante en la selección y predicción de este caracter. Este estudio intenta predecir la EA en vacas lactantes usando proxis de vacas adultas y terneras en edades tempranas, además se realiza un análisis genético de la EA en ambos grupos, vacas y terneras.

Se estimaron y analizaron cuatro rasgos relacionados con la EA. En vacas: ingesta de materia seca (IMS), eficiencia de conversión de alimento (ECA), consumo residual (CR) y beneficio económico bruto (BB). En terneras: IMS, ECA, CR y ganancia de peso diario (GPD). Para los análisis se utilizaron dos grupos, 1558 registros correspondientes a 104 vacas en producción, y 1141 registros correspondientes a 63 terneras de edades tempranas. Se realizó un análisis de predicción de los fenotipos de los caracteres relacionados con EA usando regresión lineal (LM) y tres algoritmos de aprendizaje automático: K-vecinos más cercanos (K-NN), redes neuronales (NNET), y árboles predictores (Bagging). La predicción se evaluó usando validación cruzada. Se estimaron los parámetros genéticos de cada uno de los caracteres usando inferencia bayesiana (heredabilidad y correlación genética). Se realizó un estudio de asociación del genoma completo a través de una regresión lineal frecuentista (GWAS), LASSO bayesiano, y bosques aleatorios. Finalmente se realizó una predicción genómica usando edades tempranas como proxis y GBLUP como modelo predictor, la precisión se evaluó con correlación genética y error cuadrático medio entre el valor de cría genómico estimado (GEBV) y el fenotipo corregido.

Se obtuvieron correlaciones fenotípicas altas entre IMS y CR tanto en vacas como en terneras, mientras ECA y BB mostraron correlaciones altas entre sí. En la predicción fenotípica, el método estadístico con mayor precisión para IMS fue Bagging, para ECA fue LM y para BB fue NNET, mientras que para CR las precisiones fueron muy bajas con todos los modelos. Las estimas de heredabilidad fueron altas para todos los caracteres, sin embargo, las correlaciones

genéticas entre los caracteres de vacas y terneras fueron bajas. En los análisis de asociación se detectaron algunas regiones genómicas asociadas simultáneamente a varios caracteres de EA, sobre todo entre IMS y CR, y entre ECA y BB (GDP en terneras). En vacas se observó una mayor coincidencia de regiones comunes entre ambos métodos paramétricos (GWAS y LASSO). Finalmente, las predicciones genéticas tuvieron precisiones muy bajas. Estos resultados sugieren que la EA a edades tempranas está controlada por diferentes genes que en la EA en vacas en lactación. Debido a que el poder estadístico del tamaño de la muestra es muy bajo (<10%), no se pueden asumir conclusiones claras en los análisis genómicos.

Palabras Clave:

Eficiencia Alimentaria, GWAS, predicción genómica, ingesta de materia seca, selección genómica.

ABSTRACT

The growing demand for food due to the increase in population requires a more efficient and sustainable cattle industry. In dairy cattle, 50-70% of production costs are due to feeding, so it is necessary to optimize processes and select animals with high production and low intake, that is, high feed efficiency (FE). FE is an expensive and difficult trait to measure; therefore, genetics plays an important role in the selection and prediction of this trait. This study tries to predict FE in lactating cows using proxies from adult cows and calves at early ages. Besides, genetic analysis of FE was performed in both calves and cows.

Four traits associated with FE were estimated and analysed. In cows: dry matter intake (DMI), feed conversion efficiency (FCE), residual feed intake (RFI) and return over feed cost (ROFC). In calves: DMI, FCE, RFI and average daily gain (ADG). Two groups were used for the analyses, 1558 records from 104 Holstein cows in production, and 1141 records from 63 Holstein calves of early ages. Prediction analysis of FE-phenotypes was performed using the statistical linear regression model (LM) and three machine learning algorithms: k-nearest neighbours (K-NN), neural networks (NNET), and predictive trees (Bagging). The prediction was evaluated using cross-validation. The genetic parameters of FE traits were estimated using Bayesian inference (heritability and genetic correlation). Genome-wide association studies were performed using a frequentist linear regression (GWAS), Bayesian LASSO, and random forest. Finally, genomic predictions in cows were development using early ages as proxies and GBLUP as predictor model, the accuracies were evaluated with genetic correlation and mean square error between the estimated genomic breeding value (GEBV) and the corrected phenotype.

High phenotypic correlations were obtained between DMI and RFI in both cows and calves, whereas FCE and ROFC showed to be highly correlated. In phenotypic prediction, the statistical method with the highest accuracy for DMI was the Bagging method, for FCE the LM method and for ROFC the NNET method, while RFI presented very low precision with all models. Heritability estimates were high for all traits, however, genetic correlations of FE traits between cows and calves were low. Some genomic regions associated simultaneously with many FE traits were detected, especially between DMI and RFI and between FCE and ROFC (ADG in calves). A greater coincidence of common regions detected between both parametric

methods was observed. Finally, the genetic predictions had very low accuracy. These results suggest that FE at earlies ages is controlled by different genes than FE in lactating cows. The statistical power of the sample size is very low (<10%), then, clear conclusions cannot be assumed in genomic analyses.

Keywords:

Feed Efficiency, GWAS, genomic prediction, dry matter intake, genomic selection.

RESUM

La demanda creixent d'aliment degut a l'augment de la població requereix d'una ramaderia més eficient i sostenible. En la ramaderia lletera el 50-70% dels costos de producció es deuen a l'alimentació, per la qual cosa és necessari optimitzar els processos i seleccionar animals amb alt nivell productiu i baix consum, és a dir, amb alta eficiència alimentosa (EA). No totes les granges poden estimar EA en el seu ramat perquè el seu mesurament és costós; per això, la genètica juga un paper important en la selecció i predicció d'aquest caràcter. Aquest estudi intenta predir l'EA en vaques lactants usant proxis de vaques adultes i vedelles en edats primerenques, a més es realitza una anàlisi genètica de l'EA en tots dos grups, vaques i vedelles.

Es van estimar i van analitzar quatre trets relacionats amb l'EA. En vaques: ingesta de matèria seca (IMS), eficiència de conversió d'aliment (ECA), consum residual (CR) i benefici econòmic brut (BB). En vedelles: IMS, ECA, CR i guany de pes diari (GPD). Per a les anàlisis es van utilitzar dos grups, 1558 registres corresponents a 104 vaques en producció, i 1141 registres corresponents a 63 vedelles d'edats primerenques. Es va realitzar una anàlisi de predicció dels fenotips dels caràcters relacionats amb EA usant regressió lineal (LM) i tres algorismes d'aprenentatge automàtic: K-veïns més pròxims (K-NN), xarxes neuronals (NNET) i arbres predictors (Bagging). Es van estimar els paràmetres genètics de cadascun dels caràcters usant inferència bayesiana (heredabilidad i correlació genètica). La predicció es va avaluar usant validació creuada. Es va realitzar un estudi d'associació del genoma complet (GWAS) a través d'una regressió lineal freqüentista, LASSO bayesià, i boscos aleatoris. Finalment es va realitzar una predicció genòmica usant edats primerenques com proxis i GBLUP com a model predictor, la precisió es va avaluar amb correlació genètica i error quadràtic mitjà entre el valor de cria genómico estimat (GEBV) i el fenotip corregit.

Es van obtenir correlacions fenotípicas altes entre la IMS i CR tant en vaques com en vedelles, mentre ECA i BB van mostrar correlacions altes entre si. En la predicció fenotípica, el mètode estadístic amb menor error quadràtic per a IMS va ser el mètode de Bagging, per a ECA el mètode de LM i per a BB el mètode de NNET; mentre que CR presente precisions molt baixes. Les estimes de heredabilidad van ser altes per a tots els caràcters; no obstant això les correlacions genètiques entre els caràcters de vaques i vedelles van ser baixes. Es van detectar algunes regions genòmiques associades simultàniament a diversos caràcters d'EA, sobretot

entre IMS i CR, i entre ECA i BB (GDP en vedelles). En vaques es va observar una major coincidència de regions comunes detectats entre tots dos mètodes paramètrics. Finalment, les prediccions genètiques van tenir precisions molt baixes. Aquests resultats suggereixen que l'EA a edats primerenques està controlada per diferents gens que en l'EA en vaques en lactació. Pel fet que el poder estadístic de la grandària de la mostra és molt baix (<10%), no es poden assumir conclusions clares en les anàlisis genòmiques.

Paraules Clau:

Eficiència Alimentària, GWAS, predicció genòmica, ingesta de matèria seca, selecció genòmica.

1. INTRODUCTION

1. INTRODUCTION

1.1. Brief overview

Currently, dairy cattle faces new challenges to satisfy market demands such as maintaining profitability and being environmentally friendly. Over time, the cost of raw materials, especially soybeans and corn have increased, which causes that 50-70% of the cost of production comes from feeding; this has generated through the genetic and nutrition area, an interest on improving feed efficiency (FE) that would allow maintaining profits per unit of production (Bozic et al., 2012; Bach, 2014).

Historically, FE has been estimated through several traits; most of them need to measure dry matter intake (DMI) for their calculation. DMI is expensive to measure, and therefore it is challenging to select for. Nevertheless, with the arrival of genomic selection and the variety of statistical approaches, the prediction of difficult traits to measure has become affordable; this creates opportunities to integrate these into selection indices. Many studies have been carried out on cows and heifers; however, studies that incorporate FE in calves at early ages are scarce.

According to the high relevance of genetics in current dairy production, this project aims to analyse FE in calves and adult cows. Different FE traits will be reviewed, and their prediction will be studied using statistical models. Phenotypic predictions of FE-traits will be evaluated using four predictive models. Variance components will be estimated under a Bayesian context for FE related traits in calves and cows. Then, genome wide associations analyses for FE traits will be implemented using three regression models in both calves and cows. Finally, FE traits will be predicted using genome wide information and early life predictors.

1.2. Feed efficiency

Feed efficiency (FE) is the ability of an animal to convert the nutrients from food into production units, in dairy cattle is a measure of the ability from the cow or calf-heifer to convert nutrients from the food intake into milk or growth, respectively (Connor, 2015). FE is crucial in the farm because it increases the economic income, and the main limitation is that these traits are expensive to measure (Pryce et al., 2015). Several traits and equations have been established

to measure FE, and the convenience of each trait varies according to the type of animal, production system and research objectives. The main traits are expressed as "ratio traits" or "residual traits " (Pryce et al., 2014b). The most used ones will be described next.

1.3. Measuring feed efficiency

1.3.1. Dry matter intake

Dry matter intake (DMI) is not a direct trait to measure FE, but it is an indispensable component to calculate FE. For this reason, DMI will be studied as a FE-traits. The National Research Council (2001) proposed to calculate dry matter intake as follow.

$$\text{DMI (Kg/d)} = (0.372 * \text{FCM4} + 0.0968 * \text{BW}^{0.75}) * \left(1 - e^{(0.192*(\text{WOL}+3.67))}\right)$$

where FCM4 is 4% fat corrected milk, BW is body weight, and WOL is week of lactation; $1 - e^{(0.192*(\text{WOL}+3.67))}$ is an adjustment term for depressed DMI in early lactating. One of the limitations when using this formula is that it tends to slightly overestimate the intake.

Another method is to weigh the dry matter that is supplied to the animal and the food is restored every time it eats, in this way the actual intake of dry matter can be determined (Bach, 2005). There are several ways to measure the dry matter content of the ration. In the farm an easy way to do it is by placing a sample (100 grams) in the microwave and heating it until it loses all the water, weigh the matter without water and draw the percentage ratio of dry matter (Bach, 2005).

1.3.2. Ratio traits

They can be described as the main relationship between two traits that reflects their proportion, The most popular ones are: through "Feed conversion ratio (FCR)" or "Feed conversion efficiency (FCE)" (Berry and Crowley, 2013). These traits have been extensively used in livestock species on the farm. (Beever and Doyle, 2007; Cottle and Kahn, 2014).

1.3.2.1. Feed conversion ratio

The feed conversion ratio (FCR) is a measure to quantify how efficient an animal is in transforming the feed it consumes into the units of production. It is calculated by dividing the

dry matter intake (DMI) for production unit either in growth or milk yield (Berry and Crowley, 2013). In growing calves, it can be calculated as:

$$\mathrm{FCR} = \frac{\mathrm{DMI}}{\mathrm{ADG}} = \frac{\sum_{d=0}^{lengh} \mathrm{DMI_{kg/d}}}{\frac{\mathrm{weight_{end(kg)}} - \mathrm{weight_{start(kg)}}}{\mathrm{lenght_d}}}$$

where, DMI is the dry matter intake on the days of the experiment and ADG means the average daily weight gain (Khansefid, 2016).

In lactating cows, FCR is generally estimated using the weight of protein and fat (WPF), that is calculated thought the average dairy milk (ADM) adjusted by the composition of fat and protein (Hall, 2011).

$$\mathrm{FCR} = \frac{\mathrm{DMI}}{\mathrm{WPF}} = \frac{\sum_{d=0}^{lengh} \mathrm{DMI_{kg/d}}}{\mathrm{ADM_{kg/d}} * \left(\frac{\mathrm{fat_{\%}} + \mathrm{prot_{\%}}}{100}\right)}$$

For interpretation we assume that animals with a low FCR have a higher efficiency and animals with a high FCR have a lower efficiency, being FCR values of 13 poor and FCR values of 8 very good (Hall, 2011).

1.3.2.2. Feed conversion efficiency

Feed conversion efficiency (FCE) is very similar to FCR. They differ in their interpretation. The higher the FCE value, the more efficient the animal is, the calculation in growing calves or heifers is equal to:

$$\mathrm{FCE} = \frac{\mathrm{ADG}}{\mathrm{DMI}} = \frac{\frac{\mathrm{weight_{end(kg)}} - \mathrm{weight_{start(kg)}}}{\mathrm{lenght_d}}}{\sum_{d=0}^{lengh} \mathrm{DMI_{kg/d}}}$$

In lactating cows, the formula can be adjusted using corrected energy milk (ECM) that refers to kg of milk of standardized composition for protein and fat concentrations and dividing this for DMI (Tyrrell and Reid, 1965; Beever and Doyle, 2007):

$$\mathrm{FCE} = \frac{\mathrm{ECM}}{\mathrm{DMI}} = \frac{\mathrm{ADM_{kg/d}} * \frac{(383 * \mathrm{fat\%} + 242 * \mathrm{protein\%} + 783.2)}{3140}}{\sum_{d=0}^{lengh} \mathrm{DMI_{kg/d}}}$$

1.3.3. Residual traits

Residual traits measure the observed intake values minus the expected intake values given productivity. They are usually calculated through a linear regression from the feed or energy intake, where the error corresponds to the residual trait. The most common is the residual feed intake (RFI) and residual energy intake (REI). The difference between them is that RFI uses DMI, while REI uses metabolic energy intake (MEI) for their estimations. The lower the value of the residual traits, the more efficient the animal is (Zamani et al., 2008; Pryce et al., 2014b).

1.3.3.1. Residual feed intake

Residual feed intake (RFI) was initially proposed by Koch et al. (1963), and it can be defined as the observed DMI minus the expected DMI. RFI is obtained from the residual of a linear regression over DMI as follow:

In growing calves:

$$\mathrm{DMI} = \mu + b_1\mathrm{ADG} + b_2\mathrm{MWT} + e$$

In lactating cows:

$$\mathrm{DMI} = \mu + b_1\mathrm{ADG} + b_2\mathrm{MWT} + b_3\mathrm{BSC} + b_4\mathrm{FY} + b_5\mathrm{PY} + b_6\mathrm{LY} + e$$

where DMI is the average intake level of the animal during the experiment; μ is the intersection of the model or the general mean; ADG is the average daily gain; MWT is the mid test body weight; $\mathrm{FY}, \mathrm{PY}, \mathrm{LY}$ refers to the fat, protein and lactose yielding respectively; $b_1, b_2, b_3, b_4, b_5, b_6$ are the regression coefficients; and the residual e is RFI; the measurements are expressed in units of weight (Berry and Crowley, 2013; Macdonald et al., 2014; Khansefid, 2016). The formula can be adjusted to the study population.

1.3.3.2. Residual energy intake

The residual energy intake (REI) has the same mechanism as RFI but uses values in metabolic energy, that is, all energy parameters must be considered as milk yield, pregnancy, growth, mobilization of body tissues, walking, etc. (Pryce et al., 2014a). It can be interpreted as follow:

$$\mathrm{NEI} = b_1\mathrm{NE_m} + b_2\mathrm{NE_l} + b_3\mathrm{NE_{preg}} + b_4\mathrm{BWCE} + e$$

where NEI is the net energy consumption; $\mathrm{NE_m}$, $\mathrm{NE_l}$, $\mathrm{NE_{preg}}$, BWCE are estimates of energy requirements for maintenance, lactation, pregnancy and energy changes in body weight respectively, and e is REI. The values are expressed in energy/day (Zamani et al., 2008).

1.3.4. Other traits for measure feed efficiency

Other traits to measure FE have been described or interpreted by some authors and these are subject to the study population. Milk and bodyweight was an alternative in the past to select for FE, but this is no longer a priority (Gonzalez-Recio et al., 2014). Seymour et al. (2020) describe the return over feed cost (ROFC) as a FE-trait, which is obtained from the subtraction from the price of milk (penalized by fat and protein) of the cost of cow feeding. In calves, a simple method of measuring FE could be ADG. That means FE can be adjusted to the farmer's needs and according to the database availability.

1.3.5. Advantages and limitations of each FE trait

Advantages for ratio traits

- Relatively straightforward to measure or calculate in stable systems with controlled feeding (Beever and Doyle, 2007).
- A quick measuring tool, so FCR and FCE are useful for monitoring the feeding and milk yielding (Hall, 2011; Shike, 2013).
- Easy interpretation.

Limitations for ratio traits

- FCR and FCE are correlated with weight gain, and this results in larger animals that eat more (Shike, 2013).
- Ratio traits are correlated with production (Van Arendonk et al., 1991). Genetic improvement already selects based on milk, fat and protein yield; this means, many of their genetic value is already caught in current selection index.

- In cows with the same ratio, it is not possible to distinguish their proportions from DMI/milk production; e.g. cows with low DMI and low milk production could have the same ratio as cows with high DMI and high production (Pryce et al., 2014b).

- Ratio traits do not consider significant effects on production such as lactation, days in milk, age, weight, energy expenditure or metabolic energy. They also do not take into account environmental effects such as herd, season or feed quality; so, the efficiencies could be falsely masked by other factors. (Veerkamp et al., 1995; Dechow et al., 2002; Hutjens, 2012). An alternative solution could be to correct the FCR or FCE value by these significant effects.

Advantages for residual traits

- RFI and REI take into account the environmental and significant effects for DMI and NEI, respectively; therefore, they are traits that catch the variations of the significant effects in their population (Rauw, 2009).

- The correlation with production is very weak, whereas the correlation with DMI is strong. This presents an opportunity to improve the FE (less consumption) without affecting yield (Sainz and Paulino, 2004).

- The correlation with weight is very low, making it possible to select animals with better production without increasing the weight (Rauw, 2009).

- RFI has no strong correlation with other production traits. This implies that it has an appreciable margin for improvement (Sainz and Paulino, 2004).

Limitations for residual traits

- Residual traits estimation requires linear models, which is more complex to calculate than ratio traits.

- The correlation between body condition score (BSC) and inter and intramuscular fat (InFat) is weak, so changes and movements of fatty tissue could be underestimated (Pryce et al., 2014b).

- Accuracy and reliability of residual traits are relatively low. The reference population needs to be increased, it should be as large as possible (Pryce et al., 2012, 2014b, 2015).

- A partial problem for RFI is that it has a negative value for efficient cows. This could make their interpretation difficult.
- If BSC is used in RFI estimation, it is a subjective score and could change according to by the professional interpretation that evaluates it (Pryce et al., 2014b).

Table 1. Advantages and disadvantages for feed efficiency traits.

	Advantages	**Disadvantages**
Ratio traits	Straightforward to measure in stable systems (controlled feeding).	High correlation with body weight, selecting for ratio traits results in larger animals.
	Easy calculation.	Ratio traits are correlated with production. Genetic programs already select by production.
	Easier interpretation.	Not possible to distinguish cows with high or low yield.
	Useful to evaluate FE in the farm.	Not consider significant effects on production such as age, lactation stage or days in milk.
Residual traits	**Advantages**	**Disadvantages**
	Residual traits consider associated effects on intake.	More complex to calculate than ratio traits.
	Selecting by residual traits without affecting yield is possible.	EBV reliability compared to productive traits.
	Improve the production without increasing the weight.	BCS is a subjective score.
	Is possible to use residual traits in genetic selection.	Interpretation not easy.

1.4. Genetics of feed efficiency

1.4.1. Heritability

Heritability for FE traits ranges between 0.06 and 0.56 depending on the study. Authors like Robinson and Oddy. (2004) and Ngwerume and Mao. (1992) have reported low heritabilities of 0.06 (FCR) and 0.016 (REI) in beef and dairy cattle, respectively; whereas authors such as Arthur et al. (2001b) and Veerkamp et al. (1995) have reported higher heritabilities of 0.46 (FCR) and 0.38 (RFI) in beef and dairy cattle, respectively. However, most authors agree that FE is moderately heritable and that it can be included in genetic selection indices.

Higher heritability estimates were reported in beef cattle compared to dairy cattle, probably because most studies are carried out on young animals and FE traits have less

environmental effects. The most common traits are DMI, FCR and RFI. In dairy cattle, FE target change according to age. In growing calves, the FE is measured with weight gain, whereas in lactating animals FE is measured with milk production; the most commonly studied traits are DMI, FCE and RFI.

DMI is the most important trait because other traits are calculated from it, and it has the same interpretation in beef and dairy cattle. Table 2 shows some heritability estimates for FE traits in the literature.

Table 2. Heritability with its standard error in parenthesis for feed efficiency traits in different cattle types and ages by some authors.

Trait	Reference	Cattle type- age	h^2
DMI	Robinson and Oddy. (2004)	Young beef cattle	0.27 (0.06)
DMI	Hoque et al. (2007)	Young beef cattle	0.20 (0.12)
DMI	Torres-Vázquez et al. (2018)	Young beef cattle	0.55 (0.08)
FCR	Robinson and Oddy. (2004)	Young beef cattle	0.06 (0.04)
FCR	Arthur et al. (2001b)	Young beef cattle	0.46 (0.04)
RFI	Robinson and Oddy. (2004)	Young beef cattle	0.18 (0.06)
RFI	Arthur et al. (2001a)	Young beef cattle	0.39 (0.03)
RFI	Arthur et al. (2001b)	Young beef cattle	0.43 (0.04)
RFI	Hoque et al. (2007)	Young beef cattle	0.33 (0.14)
DMI	Williams et al., (2011)	Growing dairy calves	0.17 (0.10)
DMI	Korver et al. (1991)	Lactating dairy heifers	0.56 (0.11)
DMI	Zamani et al. (2008)	Lactating dairy cows	0.12 (0.02)
FCR	Korver et al. (1991)	Growing dairy heifers	0.18 (0.08)
FCE	Van Arendonk et al. (1991)	Lactating dairy heifers	0.37 (0.14)
RFI	Williams et al., (2011)	Growing dairy calves	0.27 (0.12)
RFI	Korver et al. (1991)	Lactating dairy heifers	0.22 (0.11)
RFI	Van Arendonk et al. (1991)	Lactating dairy heifers	0.19 (0.12)
RFI	Veerkamp et al. (1995)	Lactating dairy cows	0.38 (0.15)
RFI	Pryce et al. (2015)	Lactating dairy cows (Australian)	0.20 (0.20)
RFI	Pryce et al. (2015)	Lactating dairy cows (UK and Dutch)	0.35 (0.06)
REI	Ngwerume and Mao. (1992)	Lactating dairy cows	0.016 (n/a)
REI	Zamani et al. (2008)	Lactating dairy cows	0.21 (0.02)

Where, DMI: dry matter intake; FCR: feed conversion ratio; FCE: feed conversion efficiency; RFI: residual feed intake; REI: residual energy intake; and n/a: not available.

1.4.2. Genetic correlation with another traits

Some studies have been carried out to evaluate the relationship of feed efficiency with other productive traits. Robinson and Oddy (2004), Hoque et al. (2007), and González-Recio et al. (2014) showed negative correlations between size and FE traits (RFI and FCR) in growing animals, indicating that efficient cows tend to be larger than less efficient cows.

Koch et al. (1963) and Hoque et al. (2007) showed that the correlation of weight with ratio traits (FCR-FCE) is strong; whereas Arthur et al. (2001a) and Robinson and Oddy (2004) showed that residual traits (RFI) has a weak correlation with weight (Table 3). Hence, selecting for lower FCR (efficient) would produce heavier cows that could eat more; while selecting by low RFI (efficient) would produce cows with lower DMI without an increase in weight. Some studies reported that the correlation between DMI and ratio traits is moderate or low, and the correlation between DMI and residual traits is moderate or high (Table 3).

Several studies showed that the correlation of productive traits such as milk, fat and protein with residual traits are strong whereas the correlation between productive traits and ratio traits is weaker; this could be due to FCR and FCE consider production per unit of intake while RFI and REI measure the efficiency based on the animal's feed intake and these are corrected by production factors. Gonzalez-Recio et al. (2014) showed a high correlation (0.71) between RFI and BCS while Robinson and Oddy (2004) showed that RFI-InFat correlation was moderate (0.22); this seems to indicate that BCS does not explain all the intramuscular fat mobilization. Robinson and Oddy, (2004) and Lin et al. (2013) showed that the correlation between RFI and feeding time (FT) was positive; this indicates that animals with lower RFI take less time to eat; this is also verified by Green et al. (2013).

In Angus cattle, Hegarty et al. (2007) showed a positive correlation between methane emission and RFI (0.12). This study showed that the reduction of the daily methane emission would reduce when selecting by RFI. Delgado et al. (2019) presented a relationship between RFI and the ruminal microbiota in Holstein cows. The predictive accuracy in cows could be improved using the metagenome information.

In beef cattle, Archer et al. (2002) showed a high correlation for dry matter intake (DMI) and RFI between weaned calves and adult cows (0.94 and 0.98 respectively), while the correlation with FCR was 0.20. It showed that DMI and RFI are very similar between cows

and weaned calves, whereas FCR does not. Table 3 shows some estimates found in the literature.

Table 3. Genetic correlation with its standard error in parenthesis (standard deviation in bold) between productive traits and feed efficiency in different cattle types and ages by some authors.

				Residual		Ratio	
	Trait[1]	**Author**	**Cattle type- Age**	**RFI**	**REI**	**FCE**	**FCR**
size	STAT	Gonzalez-Recio et al. (2014)	Dairy heifers	-0.50 (**0.22**)			
	MWT	Robinson and Oddy (2004)	Finished beef C.	-0.20 (0.16)			-0.62 (0.18)
	MWT	Hoque et al. (2007)	Young beef C.	-0.61 (0.30)			-0.62 (0.35)
Weight	ADG	Arthur et al. (2001a)	Young beef C.	-0.04 (0.08)			-0.62 (0.06)
	ADG	Hoque et al. (2007)	Young beef C.	-0.95 (0.08)			-0.77 (0.11)
	WG	Koch et al. (1963)	Young beef C.			0.79 (n/a)	
	WG	Robinson and Oddy (2004)	Finished beef C.	0.09 (0.20)			-0.86 (0.10)
Feed intake	DMI	Zamani et al. (2008)	Dairy cows		0.61 (n/a)		
	DMI	Gonzalez-Recio et al. (2014)	Dairy heifers	0.03 (**0.07**)			
	DMI	Lin et al. (2013)	Dairy heifers	0.45 (0.13)			
	DMI	Koch et al. (1963)	Young beef C.			0.04 (n/a)	
	DMI	Arthur et al. (2001a)	Young beef C.	0.69 (0.03)			0.31 (0.07)
	DMI	Robinson and Oddy, (2004)	Finished beef C.	0.43 (0.15)			-0.49 (0.22)
Yielding	MY	Van Arendonk et al. (1991)	Dairy heifers	0.02 (n/a)			-0.64 (n/a)
	MY	Zamani et al. (2008)	Dairy cows		-0.05 (n/a)		
	MY	Gonzalez-Recio et al. (2014)	Dairy heifers	0.07 (**0.08**)			
	FY	Gonzalez-Recio et al. (2014)	Dairy heifers	0.02 (**0.07**)			
	PY	Gonzalez-Recio et al. (2014)	Dairy heifers	0.03 (**0.07**)			
	FPCM	Van Arendonk et al. (1991)	Dairy heifers	0.02 (n/a)			-0.93 (n/a)
	FPCM	Zamani et al. (2008)	Dairy cows		0.08 (n/a)		
FT	FT	Robinson and Oddy (2004)	Finished beef C.	0.35 (0.17)			0.78 (0.16)
	FT	Lin et al. (2013)	Dairy heifers	0.27 (0.15)			
Others	BCS	Gonzalez-Recio et al. (2014)	Dairy heifers	0.71 (**0.32**)			
	InFat	Robinson and Oddy (2004)	Finished beef C.	0.22 (0.17)			0.08 (0.28)
	ClvI	Gonzalez-Recio et al. (2014)	Dairy heifers	-0.13 (**0.15**)			
	CH4	Hegarty et al. (2007)	Finished Beef C.	0.12 (n/a)			

Where, MWT: metabolic weight STAT: stature; WG: weight gain; ADG: average daily gain; DMI: Dry matter intake; MY: Milk yield; PY: protein yield; FPCM: fat protein corrected milk; ClvI: calving interval; BCS: Body condition Score; InFat: intramuscular fat; RFI: residual feed intake; FT: feeding time; Finished beef C.: Finished beef cattle; n/a: not available.

1.4.3. Genomic regions associated to feed efficiency

Several exploratory studies have been developed to investigate associated regions that contribute to understanding the phenotypic expression of FE. Some of them are shown below.

- Khansefid et al. (2017) found 6,143 genes expressed in RFI-associated muscle, liver, and blood tissue of Angus bulls and Holstein cows, and 2,343 genes associated with RFI-EGBV (estimated genetic value). This study concludes that the expression of many genes with various biological functions are associated with RFI.
- Sherman et al. (2009) showed several quantitative trait loci (QTL) associated to FE in beef cattle: 19 QTL for RFI. The most significant QTL was on BTA 3. Twelve QTL were found for FCR; the most significant one was found on chromosome 24. Finally, four QTL were found associated to DMI, of which the most significant one located on BTA 7. The closest genes were not reported.
- Yao et al. (2013) found 188 SNP surpassing the significance threshold for RFI using random forest in Holstein cows. Thirty-eight of them were located on QTLs regions associated to RFI in beef cattle by Sherman et al. (2009); these SNPs were on BTA 3, 4, 7, 11, 12, 18, 19, 23, 24 and 25, and their closest genes were LOC5309292, KLF1, REV1, AFF3, TBC1D8, COL4A12, GAS6, LOC510844, USP43, SLC47A1, LOC784682, LOC100139490, PARN, GNA12.
- Rolf et al. (2012) found 53 SNPs explaining 54.12% of the additive genetic variation (AGV) in steer breeding value for feed intake along the BTAs 11, 14, 15, 17, 19 and 21. They also found 66 SNPs explaining 62.69% of the AGV for RFI in BTAs 3, 5, 6, 12, 15, 17 and 21. The closest genes are involved in metabolic pathways, feeding and digestion functions.
- Bolormaa et al. (2011) found 75 SNPs significantly associated with RFI located in 24 different BTAs in cross beef cattle; The most significant SNPs were located on BTA 3, 5, 7, and 8.
- Salleh et al. (2017) found 70 and 19 significant differentially expressed genes (SDR) from liver tissue associated to RFI in Holstein and Jersey, respectively. These genes act in the regulation of immunity mechanisms, steroid hormone synthesis, retinol metabolism, arachidonic acid, lipids, sugars and protein metabolism, among others.
- Hou et al. (2012) identified 240 and 274 copy number variation (CNV) in cows with low and high RFI respectively. The specific genes from low RFI (efficient cows) were mainly related to the immune response, and the specific genes from high RFI (inefficient cows) were mainly involved in the cell cycle and the development of organs and bones.

1.4.4. Countries selecting by feed efficiency

The inclusion of FE in the breading goals has been limited due to difficulties on for data collection (especially DMI). With the advent genomic selection, Australia, USA, New Zealand and the Netherlands have already implemented direct selection on FE.

In 2002, the Australian beef cattle industry incorporated the net feed intake (NFI) in the BREEDPLAN EBVs (https://breedplan.une.edu.au/about/history/). In April 2015, dairy cattle in Australia incorporated RFI in the Australian Profit Ranking (APR), the project is detailed in Pryce et al. (2015) incorporates a new breeding value based on RFI. In 2016, Netherlands also incorporated a breeding value in bulls for feed intake into their selection indices (Veerkamp et al., 2014; Jong et al., 2016).

Many countries are interested in improving feed efficiency in cows. In the study by de Haas et al. (2015), feed intake data from countries (Australia, Canada, Denmark, Germany, Ireland, the Netherlands, the United Kingdom, New Zealand, and Iowa and Wisconsin in the United States) were collected to improve the accuracy of genomic estimated breeding value for dry matter intake using a common reference population.

1.4.5. Selecting by feed efficiency

FE is a heritable trait and hence genetic selection is possible. However classical selection is economically unfeasible. Due to large cost of phenotyping the alternative was by indirect selection through highly correlated traits, that are cheaper and easier to measure such as body weight or milk yield. Examples of new phenotypes are hormones, metabolites or MIR specters. Genomic selection, with a proper reference population is an efficient strategy to select for FE (Pryce et al., 2015).

1.5. Brief overview of genomic selection

Genomic selection has revolutionized the genetic improvement programs in dairy cattle. It is the process that allows estimating the breeding value using a dense panel of single nucleotide polymorphisms (SNP-chip) and use it for breeding value purposes. Genomic selection has been extensively used for traits prediction and association studies. A summary of its characteristics and development follows:

1.5.1. History

In cattle, before the GS, the best linear unbiased predictor (BLUP) model was the top reference for genetic selection that uses the inverse of the pedigree relationship matrix to estimate the animal breeding value (Henderson, 1975, 1976). In the late 1970s, the genetic markers were discovered, and later Soller and Beckmann (1983) described a possible use of these markers for breeding purposes. It consisted of more precise relationships between animals using markers intimately linked to a quantitative trait locus (QTL). The high costs of genotyping limited this technology in its period (Lourenco et al., 2017).

Marker-assisted selection (MAS) has been also popular, which consists of generating a profile with some markers associated to genes of traits of interest. The problem was that most of the productive traits are controlled by infinite genes (Fisher, 1919), so this technique was losing interest (Lourenco et al., 2017).

With the first draft of the human genome sequence in 2001, single nucleotide polymorphisms (SNPs) began to be an opportunity for genomic sciences. Meuwissen et al. (2001) proposed some methods for GS, which would take eight years to be applied. In 2009, the first bovine genome was sequenced (The Bovine Genome Sequencing and Analysis Consortium. 2009), which allowed to identify SNPs and generate commercial dense markers chips. In the same year, the first genomic evaluation was made by AGIL-USDA in Holstein and Jersey; the first genomic dairy bull named Freddie (Badger-Bluff Fanny Freddie) was evaluated, being the best genomic bull in the world which was verified three years later from his daughters. Since then, genomic selection models have been improved; the first model used was called multistep, which used multiple analyses to combine genealogical with genomic information. Few years later, single-step genomic BLUP (ssGBLUP) was developed. It combines pedigree, genotype and phenotype in one single evaluation. Genotyping in cattle rapidly increased to millions, improving the reliability of GEV from a genomic BLUP (GEBV) (Lourenco, 2017).

1.5.2. Advantages over marker assisted selection

The goal in MAS is to select genes associated with a trait or disease and use them in the breeding programs. The causal gene or genes are detected, then, genetic selection is performed. MAS works very well in traits that are controlled by a small number of genes (Mendel, 1996), such as the myostatin gene with effect on the bovine musculature (Grobet et al., 1997) or

calpain and calpastatin genes with effect on beef tenderness (Page et al., 2002). However, in complex traits (such as milk yield trait), many genes can control the trait expression (Fisher, 1919). Boyle et al. (2017) postulated an omnigenic hypothesis; this proposes that the genes associated with a complex trait could be interconnected with many genes that do not appear to be related, and that part of the heritability could be explained by the effect of these genes. So, this revelated the limitation of MAS.

Genomic selection uses panels of thousands of SNPs distributed throughout the genome. These SNPs are expected to be in linkage disequilibrium with at least one QTL. Hence, they can be used to predict the genomic estimated breeding value (GEBV) through the SNP effect on the trait (e.g. SNP-BLUP) or the genomic relationship matrix (e.g. GBLUP), the last is the most widely used in cattle. Effectiveness of genomic selection is given by the of phenotypic variance that can be explained by the SNPs (Blasco and Toro, 2014; Lourenco, 2017). Thus, the main differences between both MAS and genomic selection are:

- MAS uses few markers; whereas genomic selection uses a panel of dense marker with many SNPs in LD.
- MAS searches for specific genes associated with a trait; whereas genomic selection uses the effect of all SNPs together.

Genomic selection has been more successful, and its advantages are evident:

- Most productive traits are governed by many genes.
- Genomic selection is more accurate than MAS and have proved to increase genetic progress.
- Genomic selection is cheaper than MAS, and there are many SNPs chips on the market.

1.5.3. GWAS vs GWP

Whole-genome association study (WGAS) and genome-wide prediction (GWP) have different objectives; GWAS searches SNPs associated to quantitative traits of interest; whereas GWP tries to predict the genetic value through the variance explained by SNPs (Blasco and Toro, 2014; Lourenco, 2017). However, both methods can be used to infer associations between genomic regions and the trait of interest.

Traditional GWAS use a dense marker-panel and makes linear regressions SNP by SNP. The results are usually evaluated under the P-value criterion; the most significant SNPs (lowest

P-value) are assumed (incorrectly) to be associated with the trait under study. The representation is given by a Manhattan plot; which is a SNPs Scatter plot that shows the chromosomes on the x-axis and their P-value on the y-axis represented in $-log_{10}$ scale. GWAS has been widely questioned due to the presence of false associations and bias for variance overestimation. These problems can be reduced with large numbers of genotyped animals (Pearson and Manolio, 2008; Blasco and Toro, 2014).

In the case of WGP multiple regression models are used to analyse the effect of all SNPs together. Some methods have been developed, for instance:

- Ridge regression BLUP (RR-BLUP) or SNP-BLUP that assumes a normal distribution for SNP effects and its variance is constant (Meuwissen et al., 2001).
- Bayesian methods, such as Bayes A, Bayes B, Bayes C and Bayes Lasso, where different variances are assumed for the SNPs (González-Recio and Forni, 2011; Jiménez-Montero et al., 2013).
- GBLUP and single-step GBLUP use the BLUP methodology incorporating genomic data through the kinship matrix to estimate the genetic merit of the animals. GBLUP uses information from SNPs through the genomic relationship matrix (G) and ssGBLUP combines genomic information and pedigree to increase the accuracy in populations with deficient genomic data. In dairy cattle, the GBLUP is the most used to estimate the GEBV.
- Machine learning algorithms are relatively new and present an interesting predictive ability. Among the most studied are random forest, boosting algorithm, and Bayesian neural network (González-Recio and Forni, 2011; González-Recio et al., 2014).

Again, the lack of major genes in productive traits is a great challenge for association studies. However, under careful interpretation and in large populations, genomic associations can be a useful tool to analyse the genome in livestock.

1.5.4. Implementation

Genomic selection involves a set of steps, and they are all indispensable. The first step is to establish selection objectives; these must be clear and achievable. Second, selection of traits to be introduced in the selection indices; they must be measurable. Third, reference population establishment and genomic breeding value estimation. Four, data collection and DNA

sampling. Fifth, predicting the genomic breeding value in candidates and finally, dissemination. Reference population, prediction, and dissemination are described below.

1.5.4.1. Reference population

The reference population is a set of animals that are phenotyped and genotyped with a dense panel of markers; these associations are used to obtain the GEBV in this population or predict the GEBV in other genetically close animals, these associations are given either directly with the SNPs, or through the genomic relationships matrix. The accuracy of genomic predictions depends on by the reference population size, the genetic relationship between the reference and prediction populations, and the trait heritability (Goddard and Hayes, 2007).

1.5.4.2. Prediction

The genomic value from animals can be estimated and predicted through several alternatives, but the most used in dairy cattle is the GBLUP that uses the traditional BLUP with the genomic relationship matrix instead of the genealogy to estimate the genetic merit of individuals. GBLUP allows to increase the accuracy between 20-50% concerning the BLUP. It can be solved by classical statistic or within a Bayesian approach (Clark and Van Der Werf, 2013); in the last case, effects are assumed to have random distributions, and computationally it uses the Gibbs sampling algorithm from Markov chain Monte Carlo (MCMC) family to obtain a sequence of samplings approximated to a specified multivariate probability distribution (Blasco, 2017).

1.5.4.3. Dissemination

Farms must use GEBV of their animals, and animals with better GEBV must be selected and prioritized on the farm. The most effective method is through disseminating the semen of the bulls by artificial insemination. Embryo transfer is also a tool with high impact.

Genomic selection has a significant impact on dairy cattle. The generation interval has been reduced in half and selection response has increased in low heritable traits and selection for expensive traits to measure is possible.

1.6. Research approach

Considering the reviewed background and the scarce studies in calves at early ages, the following hypothesis has been proposed: It is possible to predict the FE of the adult cow by measuring FE when it is a calf. The genes that control the FE in calves are the same or close to the genes in cows. For this reason, this thesis analyses the FE in calves and compares it with the FE in adult cows using genomic-wide associations and genomic prediction assays. Besides, phenotypic prediction studies of FE are carried out to find ways for reducing costs in estimating this trait. This study uses parametric and non-parametric regression models, with classical and Bayesian statistic to analyse different strategies.

2. OBJETIVES

2. OBJECTIVES

2.1. Main objective

- Predict feed efficiency in Holstein cows from proxies in early and late life-stage.

2.2. Specific objectives

- Predict FE-traits from proxies in cows.
- Predict FE traits from early life proxies.
- Analyse genomic regions involved in FE of calves and cows.

3. MATERIALS AND METHODS

3. MATERIALS AND METHODS

3.1. Data

Phenotypic data:

Data in this study were collected from two locations and two sets of animals: calves and adult cows. Calves were fed ad libitum during two periods. In the first period, intake from 26 calves and their weights were recorded during 60 days. In the second period, 37 calves were monitorized under the same conditions, during 45 days. A total of 3225 individual phenotypic records were obtained, with animal identification, dry matter intake and daily weight.

Seventy adult Holstein cows were monitorized during 15 days in BLANCA from the Pyrenees farm located in Lleida, Spain. Phenotypic data included the cow ID, sire, birth date, calving date, dry matter intake, lactation number, days in milk, milk production, protein and fat percentage, body weight and age. A total of 1063 measures were obtained. Later, 480 measures of dry matter intake from 32 cows were added for the genome association studies and genomic prediction for this trait.

Genotypic data:

Thirty out of the 63 calves were genotyped with the Affymetrix Axiom Bovine Genotyping Array 60K (60914 SNPs) and then, imputed to Illumina HD Bovine SNP chip with BEAGLE software (Browning and Browning, 2008) using sequences from the 1000 bulls genome reference population (www.1000bullgenomes.com). All cows were genotyped with the Illumina EURO12K SNP chip and imputed to Illumina Bovine 50k SNP chip (54,609 SNPs) with BEAGLE software (Browning and Browning, 2008) using the Eurogenomics reference population (www.eurogenomics.com) provided by CONAFE. Finally, the SNPs in common in both platforms (35300 SNPs) were selected using R environment.

3.2. Traits

3.2.1. Calves

Four traits related to feed efficiency (FE) were studied:

1) The average daily gain (kg/day) during the sampling period (ADG).

2) The mean dry matter intake (kg/day) during the sampling period (DMI_{calf}).

3) Feed conversion efficiency (FCE_{calf}), estimated with the following formula.

$$\mathrm{FCE}_{calf} = \frac{\mathrm{ADG}}{\mathrm{DMI}_{calf}}$$

Higher values indicate more efficient animals and lower values less efficient animals.

4) Residual feed intake (RFI_{calf}), this was defined as the difference between actual and estimated feed intake, it was calculated using a linear model on DMI_{calf} as follows:

$$\mathrm{DMI}_{calf} = \mu + b_1\,\mathrm{PERIOD} + b_2\mathrm{ADG} + \mathrm{RFI}_{calf}$$

where, DMI_{calf} was the mean daily dry matter intake during the period, μ was the intercept; b_1 and b_2 were partial regression coefficients; PERIOD was a categorical trait of the period sampling, it has two levels; RFI_{calf} was the residual term, it considers the lower values for the more efficient animals and the highest values for the less efficient animals (Pryce et al., 2014b, 2015). The linear model was implemented with the glm function of the R Stats package by R Core Team and contributors worldwide (2018).

3.2.2. Cows

In cows, four traits associated with FE were studied:

1) The average daily DMI (kg/d) during the sampling period (DMI_{cow}).
2) Feed conversion efficiency (FCE_{cow}) estimated as:

$$\mathrm{FCE}_{cow} = \frac{\mathrm{ECM}}{\mathrm{DMI}_{cow}}$$

where, ECM was energy corrected milk, that refers to kg of milk of standardized composition with respect to protein and fat concentrations (Beever and Doyle, 2007), calculated as: $\mathrm{ECM} = \mathrm{MILK} * \frac{(383*\%\,\mathrm{FAT}+242*\%\,\mathrm{PROT}+783.2)}{3140}$. Higher values belong to more efficient animals whereas lower values indicate less efficient animals.

3) Residual feed intake in cows (RFI_{cow}) was calculated using a linear model on DMI_{cow}. The variables were selected according the criteria of t -value and the regression coefficient, as:

$$\mathrm{DMI}_{cow} = \mu + b_1\,\mathrm{LACT}_{cat} + b_2\,\mathrm{DIM}_{cat} + b_3\,\mathrm{MILK} + b_4\,\%\mathrm{FAT} + b_5\,\%\mathrm{PROT} + b_6\,\mathrm{BW}^{0.75} + b_7\,(\mathrm{BW}^{0.75})^2 + b_8\,\mathrm{AGE} + b_9\,\mathrm{AGE}^2 + \mathrm{RFI}_{cow}$$

where, DMI_{cow} was the mean daily dry matter intake collected from the study; μ was the intercept; $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$ and b_9 were partial regression coefficients; LACT_{cat} was a categorical trait corresponding to the lactation, with two levels: first lactation and second or subsequent lactations. DIM_{cat} was a categorical trait corresponding to the lactation period, with two levels (≤ 90 days in milk and > 90 days in milk). MILK was the average milk production (kg/d) during the trial period; %FAT and %PROT were mean fat an protein percentage; BW means body weight in the sampling period; AGE and AGE^2 were linear and quadratic traits for the mean age (in days) during the sampling period; RFI_{cow} was the residual term. Lower values indicate larger feed efficiency (Pryce et al., 2014b, 2015). The linear model was implemented with the glm function of the R Stats package by R Core Team and contributors worldwide (2018), and the AIC function of the MASS package by Ripley et al. (2019) was used to select the variables.

4) Return over feed cost (ROFC), which represents the gross income from the milk price by subtracting the feeding cost, it was calculated as follow:

$$\mathrm{ROFC} = \mathrm{MILK}_{price€} - \mathrm{FEED}_{cost€}$$

where, $\mathrm{MILK}_{price_€}$ was the price of milk adjusted by fat and protein penalties following (Charfeddine and Pérez-Cabal, 2019): $\mathrm{MILK}_{price_€} = \mathrm{MILK}_{\mathrm{kg}} * \left[0.31_€ + \left((\%\,\mathrm{Fat} - 3.7) * 0.030_€\right) + \left((\%\,\mathrm{Prot} - 3.1) * 0.040_€\right)\right]$. The $\mathrm{FEED}_{cost_€}$ was the feeding cost which was referenced by CONAFE (2020). The formula was: $\mathrm{FEED}_{cost_€} = \mathrm{DMI}_{cow} * 0.18_€$. Higher values indicate more efficient animals.

Density plots for productive and FE traits are shown in Annex 1 and Annex 2.

3.2.3. Phenotypic correlation

Phenotypic correlations were correlated in an exploratory analysis as follow:

$$r_{y1y2} = \frac{Cov_{y1y2}}{\sigma_{y1}\sigma_{y2}}$$

where, Cov_{y1y2} was the covariance between traits y_1 and y_2; σ_{y1} and σ_{y2} were the variances of the traits. For these correlations, the variables were corrected by systematic effects.

3.3. Phenotypic prediction

Phenotypic prediction of FE-related traits was performed using a set of covariates in cows. The traits were: DMI, FCE, RFI and ROFC. The covariates were selected according to their association with FE; for this, a linear regression on DMI (it was used as reference trait) was performed with all available phenotypic covariates. The covariates with the lowest t-value and the highest regression coefficient were selected. The package used were the glm function of the R Stats package by R Core Team and contributors worldwide (2018) and the AIC function of the MASS package of Ripley et al. (2019). The set of covariates included: $LACT_{cat}$, DIM_{cat}, MILK, %FAT, %PROT, $BW^{0.75}$, $(BW^{0.75})^2$, AGE, AGE^2.

3.3.1. Prediction models.

Four different statistical regression models were used in the prediction analysis, the classic parametric model and three non-parametric models of machine learning (detailed below). The traits used into the set of phenotypic variables were rescaled between 0 and 1 using the rescale function from the scales package by Wickham and Seidel (2019), the reason was the non-parametric models predict better when the variables were standardized, this has been corroborated with previous training predictions, in the parametric model the results of prediction were the same with variables rescaled and not rescaled.

Let $\mathbf{y} = \{y_i\}$ be a vector of a FE phenotypes for *n* cows; $\mathbf{X}$ is a *p* x *n* vectors of phenotypic traits: $LACT_{cat}$, DIM_{cat}, MILK, % FAT, % PROT, $BW^{0.75}$, $(BW^{0.75})^2$, AGE, AGE^2 in cows. The regression models are shown below:

1) **Linear regression (LM):** the classical model used a multiple linear regression as follow:

$$\boldsymbol{y_i} = \mu + \sum_{p=1}^{P} b_p \mathbf{X}_{pi} + e_i$$

where μ was the intercept; b_p was the regression coefficient of the features X_p; e_i was the error.

2) **k-Nearest Neighbors (K-NN):** this nonparametric regression is based on distances from *k* closest variables, this model was inspired and detailed by Fix and Hodges (1989). The general formula is:

$$\widehat{\boldsymbol{y}}_j = \frac{1}{k} \sum_{i=1}^{k} f(y_i)$$

where $\widehat{\boldsymbol{y}}_\mathrm{j}$ was predicted FE for an animal *j*; $f(\mathrm{y_i})$ was a function that select $\mathrm{y_i}$ values from the training set based on Euclidean distance with formula: $\mathrm{d}(\mathrm{X}_i, \mathrm{X}_j) = \sqrt{\sum_{p=1}^{P}(\mathrm{X}_{pi} - \mathrm{X}_{pj})^2}$; where $\mathrm{d}(\mathrm{X}_i, \mathrm{X}_j)$ was the Euclidean distance between X_i and X_j, and k was the number of nearest Euclidean distances (nearest neighbors). In order to define k, previous training iterations were computed with different values for k; the one with least mean squared error (MSE) between real and predicted value was selected. In this study, the selected values for k were 4, 6, 8, and 4 for DMI, FCE, RFI, and ROFC respectively.

3) **Neural Networks (NNET):** this algorithm is a set of functions known as neural networks, widely detailed by Jorge Matich (2001), the formula was given by:

$$\hat{\boldsymbol{f}}(\boldsymbol{x}) = a\left(\sum_i w_i g(x)\right)$$

where $\hat{\boldsymbol{f}}(\boldsymbol{x})$ was a neuron function; $w_i g_i(x)$ was the weight and function from another neuron and a was an activation function for $\hat{\boldsymbol{f}}(\boldsymbol{x})$. The machine learning works under the backpropagation algorithm developed by Riedmiller and Braun (1993). This model was built from one input, six hidden and one output neural layers and converge with 150 iterations on average. The structure of the neural layers in this model can be seen in Annex 3.

4) **Bagging**: This model generates T pseudo-training sets by bootstrapping from the original training sample, which reduces the variance and helps to avoid the overfitting, this is described by Breiman (1996). The algorithm builds *T* decision trees and average them as:

$$\hat{\boldsymbol{y}} = \mu + \sum_{t=1}^{T} c_t h_t(r; \mathbf{X})$$

where $\hat{\boldsymbol{y}}$ was the predicted phenotype; c_t was a factor that averages the regression trees; *T* was the number of trees in the forest, where each tree was built through bootstrap aggregation; $h_t(r; \mathbf{X})$ was a Bagging predictor tree, in which **X** contain the covariates. The loss function of the model was evaluated by MSE.

Table 4. Statistical models for phenotypic prediction of FE

Regression Model	Formula	Software
Fitting Linear Model	$\boldsymbol{y_i} = \mu + \sum_{p=1}^{P} b_p \mathbf{X}_{pi} + e_i$	R Stats by R Core Team and contributors worldwide (2018). R Package, version 3.5.2.
k-Nearest Neighbor	$\hat{\boldsymbol{y}}_j = \frac{1}{k} \sum_{i=1}^{k} f(y_i)$ $\mathrm{d}(\mathrm{X}_i, \mathrm{X}_j) = \sqrt{\sum_{p=1}^{P} (\mathrm{X}_{pi} - \mathrm{X}_{pj})^2}$	Package 'FNN' by Beygelzimer et al. (2019). R package, version 1.1.3.
Neural Networks	$\hat{\boldsymbol{f}}(\boldsymbol{x}) = a\left(\sum_i w_i g_i(x)\right)$	Fit Neural Networks by Fritsch et al. (2019). R package, version 7.3-13.
Bagging	$\hat{\boldsymbol{y}} = \mu + \sum_{t=1}^{T} c_t h_t(r; \mathbf{X})$	Ipred by Peters et al. (2019). R package, version 0.9-9.

3.3.2. Cross-validation

Cross-validation is a method used to evaluate the predictive ability of statistical methods or models. It consists of dividing the database into some partitions or generating random samplings; the model is trained with n-1 partitions (training set) and predicts on the data fold that stayed out of the training set (validation set). Then, predicted values (from the validation set) are compared with the observed value using a given metric, such as Pearson correlation or

mean square error (MSE). The same process is repeated with all the partitions (or the random samples) (Elkan, 2011). This process is performed with each model, and the correlation and MSE are evaluated. The type of CV depends on the partitions. This study used random CV (RCV). One hundred iterations per analysis were performed; where 75% of the database were randomly divided for the training set and 25% for the validation set. The MSE and correlation from real-predicted values were taken for the evaluation. The best models have lower values for MSE and higher correlations values.

With these metrics, ANOVA and t-tests (Annex 4) were also performed to determine if there were statistically significant differences between prediction models . It was evaluated for each FE-traits. The package used was the Stats package by R Core Team and contributors worldwide (2018).

3.4. Variance component estimation

Variance components estimation (VCE) and consequently, correlations and heritabilities were estimated through a bivariate model in a Bayesian context:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$$

where, **y** was a 2 x *n* vector from FE or productive traits; **b** corresponds to a vector from systematic effects (LACT_{cat} and DIM_{cat} in cows, PERIOD in calves); **g** was a 2 x *n* vector from genetic effects and *e* was a 2 x *n* vector of residual effects; **X** and **Z** were incidence matrices for the respective effects. Uniform prior distribution $(-999, 999)$, were assumed for **b**. Then, **g** was assumed to be distributed as $\mathbf{g} \sim \text{N}(0, \mathbf{G} \otimes \mathbf{V_g})$, where $\mathbf{V_g} = \begin{vmatrix} \sigma^2_{g_1} & \sigma_{g_{12}} \\ \sigma_{g_{12}} & \sigma^2_{g_2} \end{vmatrix}$, and residuals as $\mathbf{e} \sim \text{N}(0, \mathbf{I} \otimes \mathbf{V_e})$; where $\mathbf{V_e} = \begin{vmatrix} \sigma^2_{e_1} & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma^2_{e_2} \end{vmatrix}$. **G** was the genomic relationship matrix and **I** was the identity matrix. The **G** matrix was estimated with the following formula:

$$\mathbf{G} = \frac{1}{N} \sum_{snp=1}^{N} \mathbf{G}_{snp_{ij}} = \frac{1}{N} \sum_{snp=1}^{N} \frac{\left(x_{snp_i} - 2p_{snp}\right)\left(x_{snp_j} - 2p_{snp}\right)}{2p_{snp}\left(1 - p_{snp}\right)}$$

where, x_{ij} and x_{ik} were the genotype of a SNP (0, 1 or 2) of an individual *i* or *j*, *N* means the number of SNPs and p_{snp} means the allelic frequency in the population (VanRaden, 2008;

Yang et al., 2010). The Gmatrix.f90 code in the Fortran language developed by Legarra et al. (2011) was used to calculate the **G** matrix.

This study estimated genetic correlations between FE related traits in calves and cows. Cow phenotypes in calves were assumed as missing. Similarly, calf phenotypes in cows were assumed as missing.

An adapted version of the TM by Legarra et al. (2011) was used. The sampling method was Gibbs and Markov Chain Monte Carlo (MCMC) widely described by Sorensen and Gianola (2007). A total of 100 000 iterations were run and the Bur-in selected was 30 000, it was decided through the Gelman diagnosis over MCMC convergence with Coda package by Plummer et al. (2019) (Annex 5 and Annex 7).

3.4.1. Heritability

Heritability was estimated using the following formula:

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{e}}^2}$$

where, σ_u^2 was the additive variance and σ_e^2 was the residual variance.

Heritability from the different analyses were averaged as:

$$\overline{h_y^2} = \frac{1}{m}\sum_{y=1}^{m} h_y^2$$

where, m was the number of bivariate analyses for the y trait.

3.4.2. Genetic correlations

Posterior distribution for the genetic correlations were estimated. The correlations had the following formula:

$$\mathrm{r}_{\mathrm{g1,g2}} = \frac{\mathrm{Cov}_{\mathrm{g1,g2}}}{\sigma_{\mathrm{g1}}\sigma_{\mathrm{g2}}}$$

where, $\mathrm{Cov}_{\mathrm{g1,g2}}$ was the genetic covariance for trait 1 and 2 respectively. The correlations and their graphical presentations were developed using the Corrplot, R Stats and GGally

packages by Taiyun Wei et al. (2017), R Core Team and contributors worldwide (2018) and Schloerke and Crowley (2020) respectively.

3.5. Genome wide association study

Genomic regions associated with FE traits were investigated using whole genome information. Four steps were implemented: 1) Calculate the statistical power of the sample size, 2) principal component analysis of **G** matrix, 3) whole-genome association studies under three statistical models and 4) results interpretation.

3.5.1. Statistical power of the sample size

Finding a significant effect for an SNP among 35300 is like finding a needle in a haystack, and therefore genome-wide association studies require significantly large sample size to avoid spurious results. The classic significance method is determined by an alpha threshold given by a p-value ($p = \Pr(T \geq t_{\text{obs}}|H_0)$). However, the p-value in genomics has been widely criticised. One of the main problems is that p-values have a poor relationship with the strength of the evidence for a real effect in different sample sizes. Bayesian inference using the Bayes factor (strength of evidence in the data), which uses prior distribution as a representation of our prior knowledge may alleviate this problem. In this study, the power of the sample size was estimated using a Bayes factor large enough to obtain reliable significant effects (Gondro et al., 2013).

For this, some factors that affect the statistical power of the sample size were considered, these include the QTL frequencies (p and q), the heritability of the QTL ($h_Q{}^2$), the linkage disequilibrium coefficient (D) and a required Bayes factor (B) as Gondro et al. (2013):

$$B = \frac{\Pr(y|H_1)}{\Pr(y|H_0)} = \frac{\Pr(H_0|y)}{\Pr(H_1|y)} = B \times \frac{\Pr(H_0)}{\Pr(H_1)}$$

B represents the factor by which the prior odds increase to give later probabilities after observing the data. So, the calculation of the B in this study was determined as:

$$B \approx [4n^2p^3(1-p)^3]^{-1/2}\left[1 + \frac{2}{(n-3)}F\right]^{n/2}$$

where n was the total sample size and F was the classic value of F-value.

The calculation to estimate a given value of *F* (F_c) was:

$$F_c = \frac{n-3}{2}\left([4n^2p^3(1-p)^3\ B_c^2]^{1/n} - 1\right)$$

Then, the statistical power of the sample size was implemented as:

$$Pw = 1 - F(F_c; v_1, v_2, \delta)$$

where v_1 = 2 was the degrees of freedom being tested and v_2 was the degrees of freedom of the error and δ was the non-centrality parameter.

The statistical power of the size of a reliable sample is 0.80 with a Bayes factor = 20. We calculated the following statistics:

- Statistical power that explains a 5% and 1% of the variation assuming an equivalence with a Bayes factor = 20.
- The sample size required for a power of 0.8 and a Bayes factor = 20.

We assumed a favourable linkage disequilibrium between the QTL and the marker of 0.25, and values for both alleles *p* and *q* of 0.5. The estimation was done with the R LdDesing package developed by Ball (2012).

3.5.2. Accounting for the polygenic effect in the GWAS

A principal component analysis (PCA) to reduce the dimensionality of the **G** matrix was performed. It was calculated through the eigen decomposition of the correlation matrix of the **G** matrix.

$$\mathrm{PC_j} = \alpha_{1j}\mathrm{V_1} + .. + \alpha_{nj}\mathrm{V_n}$$

Where coefficients α_{ij} were the elements of the eigen vector of the jth eigenvalue; $\mathrm{V_1}$ to V_n were the variables of **G** matrix (Macciotta et al., 2010). The result was a set of PCs that explains the variance with fewer features. The PCs that explain more than 60% of the variance were selected. The R Stats and factoextra packages were used (R Core Team and contributors worldwide, 2018; Kassambara and Mundt, 2019).

3.5.3. Models

Genome-wide association studies (GWAS) with FE in cows and calves were computed under three statistical models: frequentist linear regression, Bayesian linear regression with LASSO method, and Random Forest regression. The objective of using three association models was to evaluate the recurrence of relevant SNPs between models and between FE traits.

3.5.3.1. Frequentist GWAS

The classical GWAS model implements a simple linear regression SNP by SNP as follow:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{SNP_i}a_i + \mathbf{PCC}_{60\%} + \mathbf{e}$$

where $\mathbf{b}$ was the vector from systematic effects; $\mathbf{SNP_i}$ was a vector with the SNPs genotypes (AA, Aa or aa represented as 0, 1 or 2 respectively); a_i was the allelic substitution effect from SNP_i (Aguilar et al., 2019); $\mathbf{PCC}_{60\%}$ was the PCs that explain more than 60% . The evaluation was performed through P-value test.

The significance level chosen was 0.001, this mean $P < \alpha = 0.001$ was considered significant. The linear regressions were estimated using the R Stats package.

3.5.3.2. Bayes LASSO association

This model is a combination of Bayesian regression with classic LASSO (Least Absolute Selection and Shrinkage Operator) regression developed by Tibshirani (1996). This approach shrinkages the marker effects to increase the effectiveness in specifics regions (De Los Campos et al., 2009). Bayes LASSO can be represented as:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{XF} + \mathbf{Z\beta} + \mathbf{e}$$

where $\mathbf{y}$ was a vector of phenotypes; μ was the population mean; $\mathbf{XF}$ was the systematic effects matrix composed by $\mathrm{LACT}_{cat}, \mathrm{DIM}_{cat}, \mathrm{PCC}_{60\%}$ in cows and $\mathrm{PERIOD}, \mathrm{PCC}_{60\%}$ in calves; $\widehat{\boldsymbol{\beta}}$ was the LASSO estimation and $\mathbf{Z}$ was the SNP incidence matrix (0, 1 or 2); e was the vector from residual effects with normal distribution and unknown variance ($\mathbf{e} \sim (0, \sigma_e^2)$) (Park and Casella, 2008; González-Recio and Forni, 2011).

Within a Bayesian context, a posteriori distribution by LASSO is possible when the prior has an independent and identical double exponential distribution, also known as the Laplace distribution (De Los Campos et al., 2009). The prior was proposed as follow:

$$p(\boldsymbol{\beta}|\sigma_e^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma_e^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma_e^2}}$$

where σ_e^2 was the residual variance, with a Scaled-inverse Chi-square density: $p(\sigma_e^2) = \chi^2(\sigma_e^2|\mathrm{S}_e, df_e)$ and λ was a parameter that controls the level of shrinkage of the distribution

(Park and Casella, 2008; González-Recio and Forni, 2011; Pérez and De Los Campos, 2014). In both groups, the prior density for σ_e^2 was assigned with: $df = 3$; S =0.25; and λ prior parameters: shape=0.52, rate=1e-5, value = it was adjusted by posteriori λ results (Annex 9 and Annex 10), type='random'. The computation was implemented under a Gibbs sampling with Markov chain Monte Carlo algorithm (MCMC) with 100 000 iterations, discarding the first 30 000 and drawing a value every 10 iterations. The BGLR package by Perez and De Los Campos (2018) was used. Additionally, the additive genetic variance was also estimated with the formula $\sigma^2{}_{SNP_I} = \mathrm{SNP}_{effect}{}^2(2pq)$ (Carvalho et al., 2020).

3.5.3.3. Random forest

Random Forest (RF) algorithm was proposed by Breiman (2001). RF is an ensemble learning method that creates a multitude of decision trees (forest) and outputs the mean prediction of the individual trees. This model generates T pseudo-training sets by bootstrapping from the original training sample, which reduces the variance and helps to avoid the overfitting**.** This algorithm differs from Bagging because during training the RF algorithm randomly selects a subset of available variables (SNPs) for selection in each split in the tree (Hempstalk et al., 2015). This method was selected because the genome wide analysis implies analysing many SNP variables and it has shown robust results in association studies (González-Recio and Forni, 2011; Yao et al., 2013; González-Recio et al., 2014; Hempstalk et al., 2015). The formula used was:

$$\mathbf{y} = \mu + \sum_{t=1}^{T} c_t h_t(\mathrm{r}; \mathbf{W}_\mathrm{m})$$

where $\mathbf{y}$ was a vector of phenotypes; μ was the population mean; $h_t(\mathrm{r}; \mathbf{W}_\mathrm{m})$ was each random tree; $\mathbf{W}$ was to matrix effects composed by LACT_{cat}, DIM_{cat}, $\mathrm{PCC}_{60\%}$ and SNPs genotype, and c_t was a shrinkage factor that averaged the trees.

The parameters for this model were: T = 10000 trees; a maximum branches per tree = 5000; $m \approx \sqrt{n_W}$. The Loss function was evaluated by the MSE. This analysis was performed with RanFoG Software by Gonzalez-Recio (2010). SNPs with highest importance within a quantile >0.999 were selected and classified.

Variable importance

The significance of the markers over the FE traits were measured by the importance of each SNP. The importance was a representation of the SNP influence on the prediction accuracy for the FE trait. It may be summarized as:

1) In each random sampling to construct a tree, a smaller percentage of data called out of bag (OOB) remains. After each tree was formed, the prediction accuracy of the FE trait was calculated with the tree and OOB data.
2) The values of the m^{th} SNP in the OOB were permuted and the prediction accuracy was calculated again.
3) The difference between these prediction accuracies (with original OOB and permuted OOB) were calculated.
4) This process was repeated with all the SNPs and then, the difference between the prediction accuracies was averaged over all the trees of the RF.

Finally, a value for each SNP was obtained which represent the SNP importance.

3.5.3.4. Results Interpretation

The following points were analysed for better interpretations of the results.

- For GWAS, significant SNPs with $P < \alpha = 0.001$ were considered. For Bayes LASSO, SNPs with the highest effect in a quantile> 0.999 were considered. For random forest, SNPs with the highest importance in a quantile> 0.999 were considered.
- Number of SNPs in common between traits were calculated using the VennDiagram package by Maintainer and Boutros (2018) in R environment. Where, WGAS consider

a $P < \alpha = 0.001$, while Bayes LASSO and random forest consider a quantile > 0.999 as parameters to select SNPs.

- The most significant SNPs where detected with their closest genes and their biological functions using the Ensembl organization (www.ensembl.org) and the uniport organization (www.uniprot.org). For this, GWAS select the SNPs with the lowest *p*-value, while Bayes LASSO and random forest consider a quantile > 0.9999 to select the SNPs with the highest effect and importance, respectively.
- In Bayes LASSO the additive genetic variance also was estimated, and it was represented in a Manhattan plot.

3.5.3.5. Analysis between models

Finally, a coincidence analysis of the relevant SNPs was performed to detect common regions between models. For this:

- Significant SNPs from GWAS, Bayes LASSO and random forest ($P < \alpha = 0.001$ and quantile > 0.999, respectively) were matched and then identified.
- The repeated SNPs in the three models were identified, then, the nearest genes and their biological function were searched using the Ensembl organization (www.ensembl.org) and the uniport organization (www.uniprot.org).
- Besides, common SNPs between calves and cows within the significant SNPs were tried to identify.

Graphics and SNPs identification were development with R Base, VennDiagram, and ggplot2 packages by R Core Team and contributors worldwide (2019), Maintainer and Boutros (2018), and Wickham et al. (2020) respectively.

3.6. Genome wide prediction

Genomic prediction of breeding value of FE-traits (DMI, FCE and RFI) in cows was implemented using genomic and phenotypic data from calves at early life-stage. For this, genetic markers and phenotypic data collected from calves (DMI, FCE, RFI) were used as the reference population (training set). GBLUP model was used for all traits as follow:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$$

where $\mathbf{y}$ was a vector of phenotypes (cow's phenotype were assumed as missing values); $\mathbf{1_n}$ was a vector of ones (1 x n); μ was the population mean; $\mathbf{b}$ was a vector from systematic effects (PERIOD); $\mathbf{g}$ was a vector of genomic breeding values; $\mathbf{X}$ and $\mathbf{Z}$ were respective incidence matrices (All cows were grouped at level 1 of PERIOD) and e was the residual vector. The effect distributions were previously described in VCE analyses (point 3.4). Then, the genomic breeding value was predicted as:

$$\hat{\mathrm{g}} = \left[\mathbf{Z'Z} + \mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_g^2}\right]^{-1}[\mathbf{Z'}(\mathrm{y} - \mathbf{1_n}\hat{\mu})]$$

where $\hat{\mathrm{g}}$ was the vector of GEBV $\mathbf{Z'}$ was the transpose $\mathbf{Z}$ and $\hat{\mu}$ was the estimated mean. The software used was an adapted version of the TM by Legarra et al. (2011) in a Bayesian context.

3.6.1. Accuracy of prediction

The objective of this analysis was to determine how close the predicted breeding value using in early life was to the true breeding value in cows. The true breeding value was unknown; however, the phenotypes were available, so the accuracy was evaluated using the phenotype corrected by systematic effects, the phenotypes were the same that were used by developing the BLUP; i.e. phenotypes were rescaled between 0,02 and 1. The accuracy of predictions was evaluated through two methods as follow:

1) The correlation between the GEBV and the corrected phenotype as:

$$\rho_{\hat{\mathrm{g}},\mathrm{y}} = \frac{\mathrm{Cov}_{\hat{\mathrm{g}},\mathrm{y}}}{\sigma_{\hat{\mathrm{g}}}\sigma_{\mathrm{y}}}$$

where $\hat{\mathrm{g}}$ was the predicted breeding value (GEBV) and y was the phenotype corrected by LACT_{cat} and DIM_{cat}.

2) The mean squared error (MSE) between the GEBV and the corrected phenotype as:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(\hat{\mathrm{g}}_i - \mathrm{y}_i)^2}{n}$$

where n was the number of cows used in the model.

Figure 1 show an illustration of genomic prediction and accuracy evaluation. Accuracies were developed in R environment.

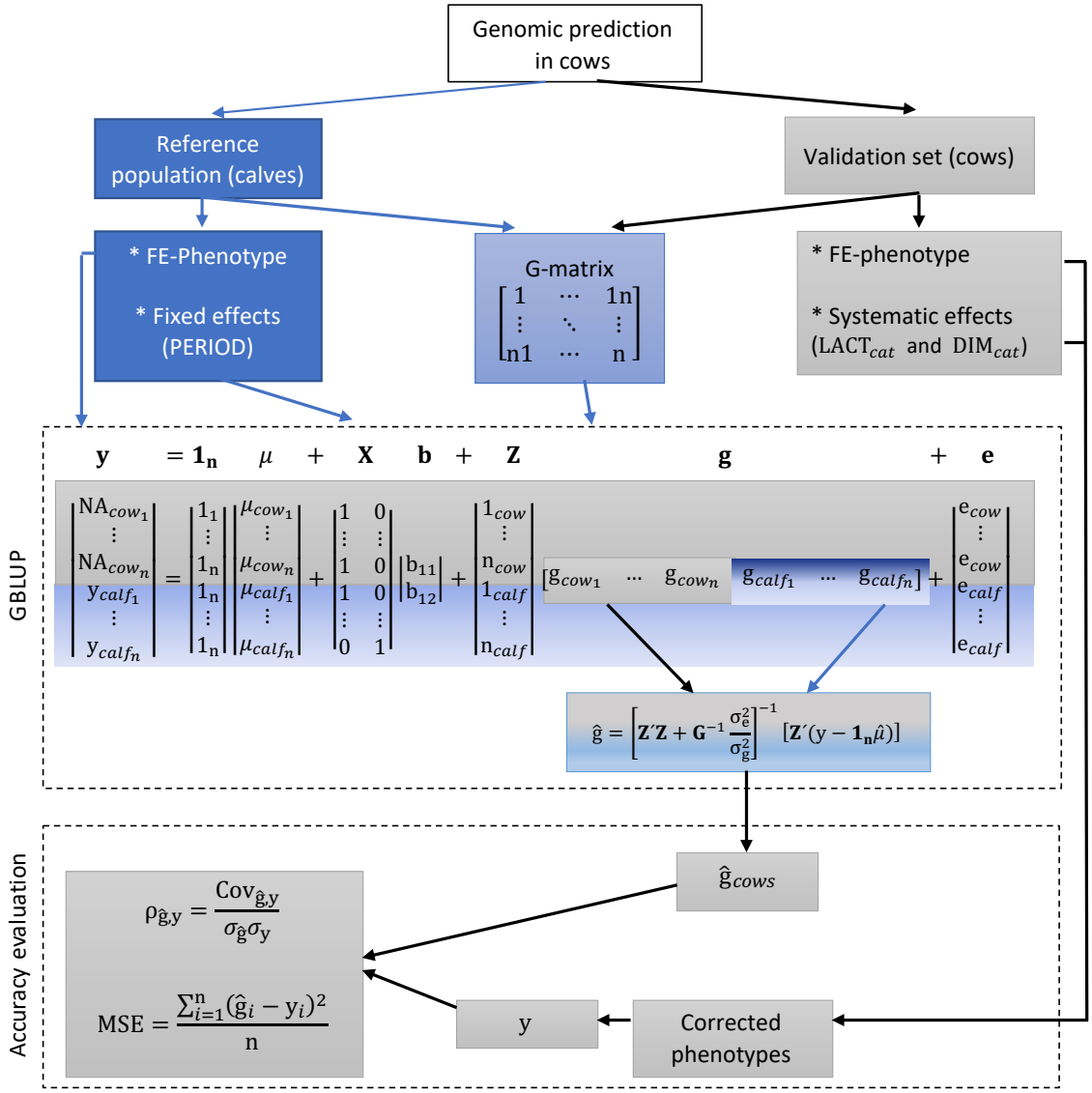


Figure 1. Summary of the prediction process and its accuracy evaluation

4. RESULTS AND DISCUSSION

4. RESULTS AND DISCUSSION

4.1. Exploratory data analysis

In calves, FE-traits showed asymmetric bimodal distributions for DMI, FCE and ADG (Annex 1), this was clearly influenced by the sampling period as in period 1 the sampling was for 60 days, while in period two the sampling was for 45 days. This significantly affects the distribution the phenotypes. RFI had a unimodal distribution because it was the residual of a linear regression on DMI corrected by PERIOD.

In cows, FE-traits distributions were slightly asymmetric for DMI, FCE and ROFC (Annex 1). This was caused by the environmental factors. The most relevant ones were lactation and days in milk. In the case of RFI it was the residual term, so it presented symmetric distribution.

Herd summary: Most of the cows were young and were between the first and third lactation, therefore, the number of cows was represented in two groups, Primiparous and multiparous. Most cows were in the second third of the lactation, so DIM cat was categorized into two groups, until day 90 and from day 91 onwards, this was divided according to milk production. Yield traits, weight and age were within the breed parameters (www.mapa.gob.es/es/ganaderia; www.conafe.com/estadisticas.aspx). Table 5 and Table 6 descriptive statistics of the data, while Annex 1 and Annex 2 show density curves for the variables on study.

Table 5. Summary from FE-traits

		DMI	FCE	RFI	ADG/ROFC
Calves	Min	0.888	0.031	-0.485	0.067
	Median	1.204	0.400	0.017	0.453
	Mean	1.710	0.299	0.000	0.383
	Max	2.789	0.588	0.286	0.691
Cows	Min	16.78	1.193	-5.013	4.992
	Median	24.11	1.596	0.047	8.512
	Mean	24.29	1.617	0.000	8.782
	Max	28.58	2.356	3.898	14.198

DMI= dry matter intake; FCE = feed conversion efficiency; RFI = residual feed intake; ADG = average daily gain; ROFC = return over feed cost. Higher values from RFI are for less efficient animals and lower values for

the more efficient animals; higher values from FCE, ADG and ROFC are from more efficient animals and lower values for less efficient animals.

Table 6. Summary of productive traits in cows

	DIM	MILK	% FAT	% PROT	ECM	BW	AGE
Min	55.00	28.08	3.008	2.841	27.05	449.4	684.0
Median	81.00	40.74	3.689	3.435	38.46	613.2	1151.0
Mean	84.86	40.89	3.709	3.411	39.31	604.9	1205.5
Max	130.00	55.50	4.639	3.862	56.09	739.2	2949.0

DIM=days in milk; MILK= milk production expressed in kg/day; %FAT= fat production expressed in percentage; %PROT= protein production expressed in percentage; ECM= energy corrected milk; BW= body weight expressed in kg; AGE= age cow expressed in days.

4.1.1. Phenotypic correlation

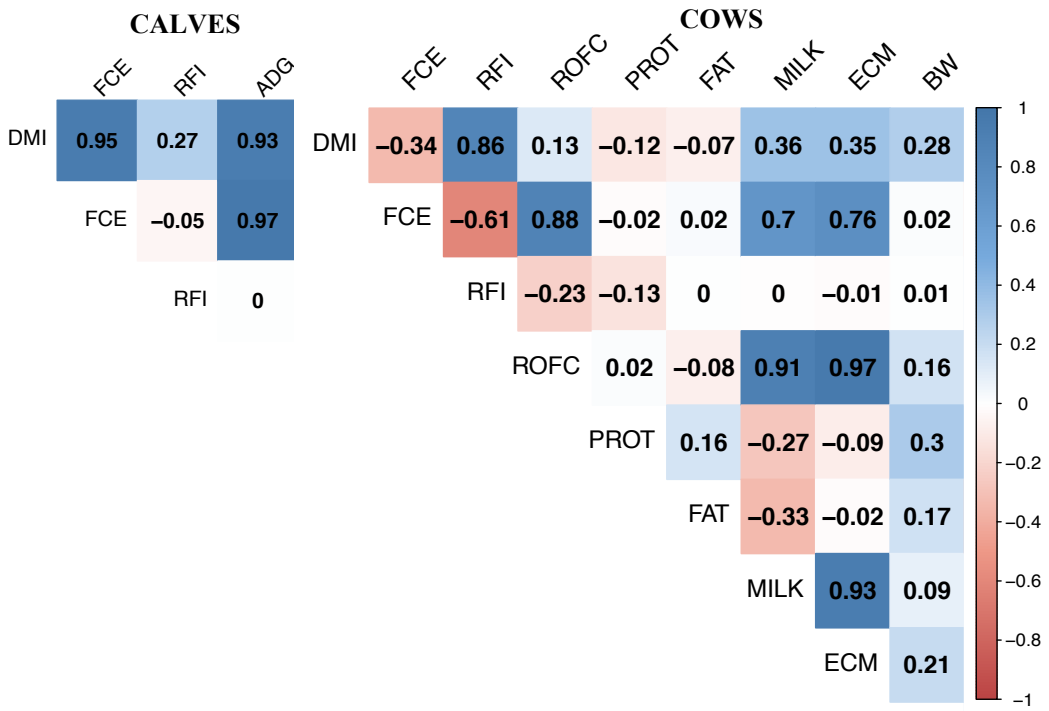


Figure 2. Phenotypic correlations for traits under study, the traits were corrected by systematic effects.

In calves, there was a strong correlation between DMI-FCE and ADG, all three traits were strongly correlated (between 0.93 and 0.97), while RFI had a low correlation with DMI (0.27) and was not correlated with any other trait.

In cows, DMI had a strong correlation with RFI (0.86), indicating that cows that eat less were more efficient. DMI was moderately correlated with milk production (0.36), ECM (0.35)

and body weight (0.28), these correlations suggest that cows with higher feed intake produce more and had more weight but showed lower efficiency from RFI point of view. RFI was not correlated with productive traits; this was because it has been adjusted in the model. FCE was strongly correlated with ROFC (0.88), which both tend to classify the same animals as efficient; furthermore, FCE and ROFC were strongly correlated with MILK and ECM. Phenotypic correlations are shown in Figure 2.

4.2. Phenotypic prediction

As detailed in materials and methods, a phenotypic prediction of FE traits was performed with four statistical models, classical linear regression and three machine learning algorithms. Then, they were evaluated with random cross validation (RCV). The objective of phenotypic predictions was to find alternatives to expensive DMI measurements to estimate FE. Therefore, DMI was excluded from the training variables in all predictions. The results of the cross-validation were the correlation (r) and the mean square error (MSE) between real and predicted values of FE-traits.

4.2.1. FE-traits comparative

The accuracy from DMI prediction was moderate; it had *r* values between 0.47 and 0.58 and MSE values between 3.53 $(kg/d)^2$ and 4.95 $(kg/d)^2$. Bagging was the model with higher accuracy followed by LM; these results were higher than those of Roseler D. K. et al. (1997), who carried out studies in 241 Holstein cows from different EEUU regions, in which the lowest MSE was equivalent to 4.7 $(kg/d)^2$. This could be due to the selected variables for the training set, the trait variance in the sample population or maybe, to the predictor model. According to the linear regression, the variables with the greatest importance (regression coefficient) to predict DMI were body weight and age, indicating that weight and age directly influence in DMI.

FCE had good accuracies, with *r* values between 0.71 and 0.85 and its MSE values were between 0.018 and 0.027; the models with the highest predictions for FCE were LM and NNET. According to the linear regression, the predictor variables with the greatest importance were Milk body weight.

RFI had lower predictive accuracy, with negatives r values between -0.31 and 0.01, and its MSE values were between 3.20 Kg^2/d and 4.74 Kg^2/d. No model was phenotypically

correlated with this trait (Figure 3), probably because RFI was already indirectly adjusted for these traits. The predictor variables with the highest regression coefficient for RFI was age. However, they had low prediction accuracy of RFI.

ROFC had high accuracies; it had r values between 0.88 and 0.99, and its MSE values were between 0.13 $euros^2$/d and 1.2 $euros^2$/d; the best predictors were LM and NNET; both had an excellent prediction capacity. Except for DMI, no references for phenotypic predictions have been found. The predictor variables with the greatest importance were MILK and body weight, which also suggests that ROFC was largely determined by the production and weight.

4.2.2. Models comparative

The model's accuracy differs between FE traits; this is shown in Figure 3. LM had shown great predictive ability in general; its highest precision was for FCE and ROFC. K-NN showed consistent accuracies.

NNET showed high predictive ability for ROFC. This is shown in Figure 4. NNET was the only deep learning method in this study, and its algorithm may be limited by the amount of data for back-propagation training. Bagging showed large abilities, its highest predictions were for DMI. It performs random sampling, generates different trees and adjust the result by MSE; this avoids overfitting and produces reliable results. In general, Bagging and LM showed larger accuracy than the other models; in this study NNET is an algorithm that could be recommended to train with more databases and more variables to exploit all its advantages. KNN presented consistent accuracies in this study, although its algorithm has been successfully tested on traits with nonlinear behaviours and categorical variables.

Table 7. Pearson correlation of RCV in phenotypic prediction of FE

		DMI		FCE		RFI		ROFC	
	Model	r	sd	r	sd	r	sd	r	sd
COWS	LM	0.547	0.174	**0.852**	0.052	-0.309	0.187	0.987	0.005
	K-NN	0.480	0.240	0.715	0.129	-0.309	0.248	0.884	0.042
	NNET	0.480	0.204	0.831	0.065	**0.010**	0.216	**0.988**	0.004
	Bagging	**0.582**	0.162	0.807	0.046	-0.047	0.255	0.943	0.022

Where **r** was the RCV results and corresponds to the mean from 100 Pearson correlations and its standard deviation. Higher values are desired, the model with the highest correlation is marked in boldface. LM= linear model; K-NN= k-nearest neighbor; NNET= neural network.

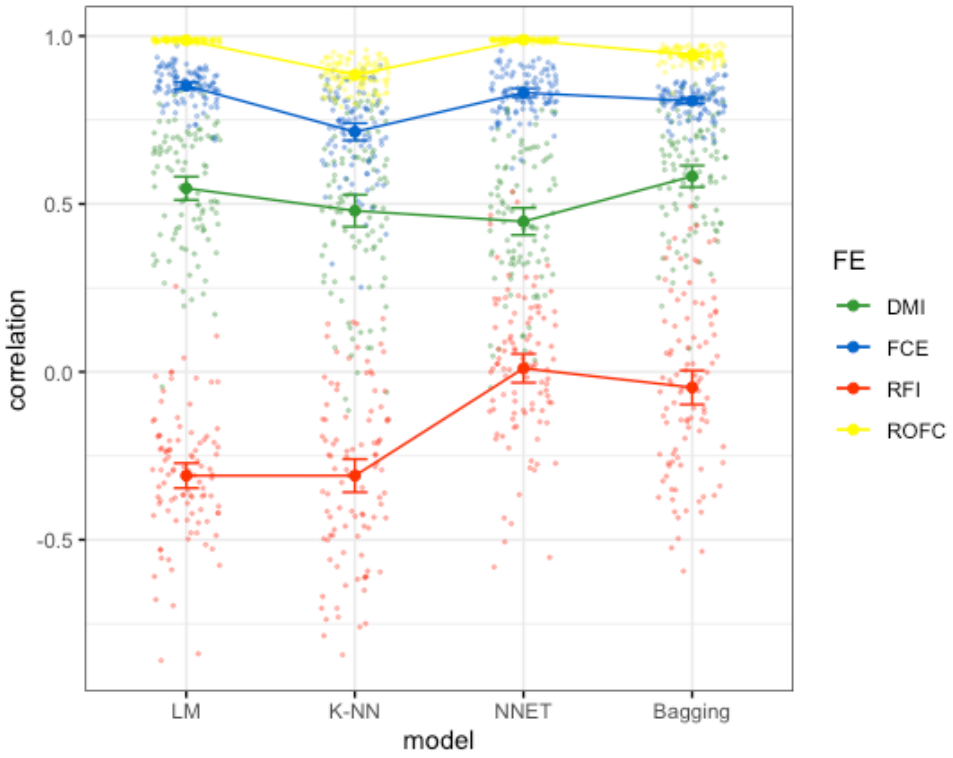


Figure 3. Correlation between the observed and predicted value from RCV, and its confidence interval. Each point represents a correlation and each colour represent a FE trait. Higher values are desirables.

Table 8. MSE and its sd of RCV in phenotypic prediction of FE

		DMI		FCE		RFI		ROFC	
	Model	MSE	sd	MSE	sd	MSE	sd	MSE	sd
COWS	LM	3.863	1.222	**0.018**	0.008	3.701	1.250	**0.126**	0.040
	K-NN	4.121	1.444	0.031	0.012	3.478	1.186	1.199	0.426
	NNET	4.948	1.750	0.021	0.009	4.742	1.374	0.131	0.049
	Bagging	**3.527**	1.097	0.023	0.009	**3.196**	0.967	0.589	0.265

MSE = mean squared error, this was the RCV results and correspond to the mean from 100 mean squared errors and its standard deviation. Lower values are desired; the model with the highest correlation is marked in boldface.

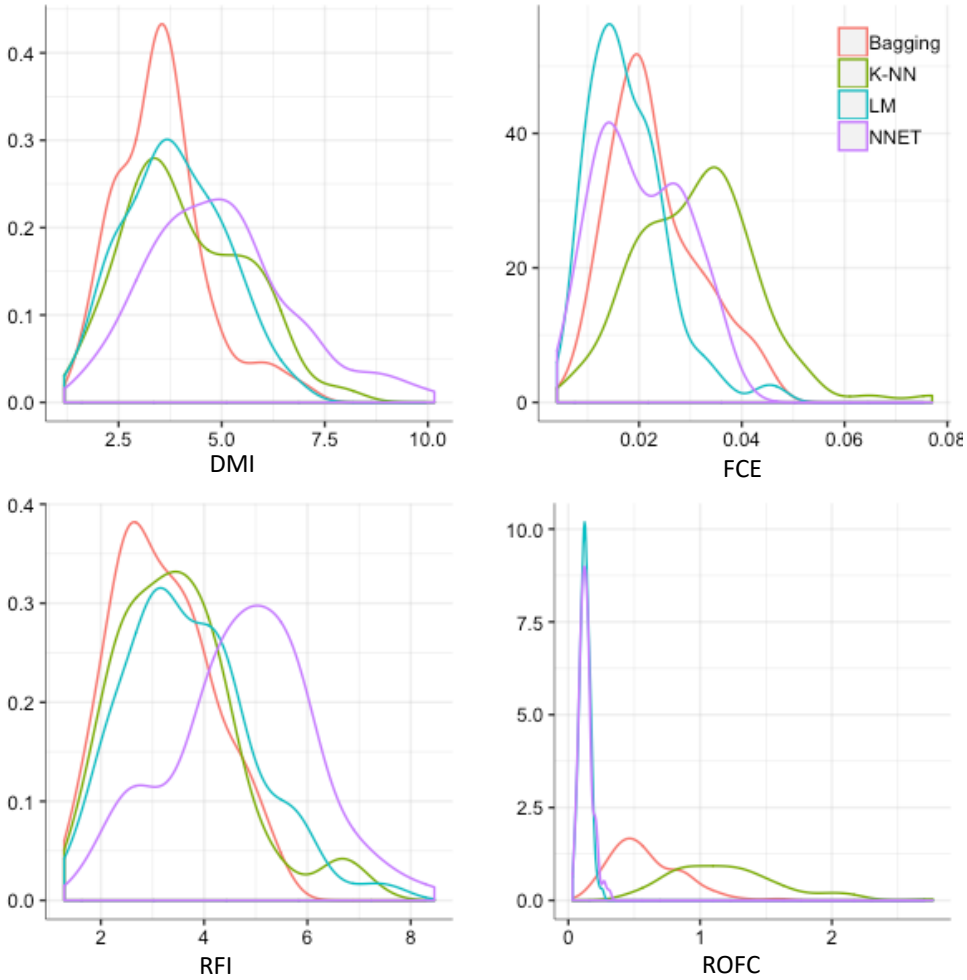


Figure 4. Density plots of MSE from RCV. The four models in each FE trait are shown and are divided by colour.

4.2.2.1. ¿Is there a statistical difference between the models?

With the results of the accuracy predictions (*r*), ANOVA tests were performed to verify if there is a statistically significant difference between prediction models. The results register a significant difference between predictive models, this was repeated for all FE traits, where the Pr (>F) for DMI, FCE, RFI and ROFC were 7.107e-07, 7.049e-32, 1.351e-31 and 3.06e-116 respectively. The t test analysis were also performed to determine the difference between prediction models (two-by-two comparisons), these results reinforce the ANOVA analysis and its results are shown in Annex 4.

4.3. Variance components analysis

4.3.1. Heritabilities

In calves, the heritabilities of FE traits were high with values of 0.50 (0.27), 0.52 (0.24), 0.50 (0.22) and 0.50 (0.26) for DMI (sd), FCE (sd), RFI (sd) and ADG (sd) respectively.

In cows, the heritabilities for FE traits were also high. Interestingly, DMI had the lower estimate (h^2 = 0.38, sd = 0.25), these results are comparable to those in Spurlock et al. (2012) and Li et al. (2016) who presented heritabilities between 0.20 and 0.40. FCE showed high heritability (h^2= 0.49; sd = 0.25) and these results were higher than those of Van Arendonk et al., (1991) and Vallimont et al. (2011), who reported heritabilities of 0.37 and 0.14 respectively. RFI presents a high heritability (h^2= 0.44, sd = 0.25), these results were higher than those presented in Table 2. ROFC was less studied trait than the previous ones, and it also presented high heritability (h^2= 0.50, sd = 0.24). The heritabilities of the productive traits were also estimated. Milk (sd), ECM (sd), and FAT (sd) had high heritability estimated with values of 0.47 (0.21), 0.46 (0.21), and 0.58 (0.23), respectively; whereas protein and BW presented lower heritabilities (0.26 (0.18) and 0.34 (0.20), respectively). Heritabilities can be seen in Figure 5 and its sd in Figure 6.

4.3.2. Genetic correlations

In calves, DMI showed high correlation with RFI (r = 0.99; sd = 0.01); as well as FCE and ADG presented a strong correlation (0.93; sd = 0.12). correlation between DMI with FCE was low (0.22; sd = 0.60) whereas DMI was moderately correlated with ADG (0.55; sd = 0.49). The correlation between RFI and FCE was weak (0.11, sd = 0.60); RFI and ADG showed moderate correlation (0.42, sd = 0.54). Correlations must be carefully interpreted because they had large sd and the sample size was low. The interpretations of traits must also be cautiously considered, RFI prioritizes feed saving while FCE prioritizes daily weight gain. No studies references have been found in calves of these ages.

In cows, DMI showed strong correlation with RFI (r = 0.85, sd = 0.25), this indicates that cows with lower feed intake tend to save more food This value agrees with those of Arthur et al. (2001a) and was higher than those of Lin et al. (2013) (Table 3). Furthermore, DMI and RFI showed positive correlations with fat, which suggest that cows with lower intake and lower RFI (higher saved food) produce a lower percentage of fat and protein; these results differ from

those by Gonzalez-Recio et al. (2014) that presented correlation values of DMI (sd) and RFI (sd) with fat of -0.11 (0.08) and 0.03 (0.07) respectively, but using DMI recorded in heifers.

RFI was negatively correlated with milk (r = -0.41; sd = 0.51) and ECM (r = -0.29; sd = 0.56), however the sd were very large, so the interpretation must be cautious. Cows with lower RFI produce more milk, and the fat and protein percentages tend to decrease (this was why the correlation of RFI with fat and protein were positive); this was in agreement with the studies by Cue et al. (1987) that showed a negative correlation of milk production with fat and protein per cent.

FCE presented a strong correlation with ROFC (r = 0.87; sd = 0.18) and both FCE and ROFC were strongly correlated with milk and ECM (Figure 5). Furthermore, FCE and ROFC were negatively correlated with fat percentage (due to the cows with high FCE and ROFC produce more milk and milk was negatively correlated with fat). ROFC showed weak correlation with body weight (r = 0.06; sd = 0.63).

The genetic correlations between cows and calves were close to zero. DMI of calves with DMI of cows had r = 0.14 with a large sd (0.70); FCE of calves with FCE of cows had r = -0.03 (sd = 0.72). Calf RFI showed r = 0.04 (sd = 0.69) with cow RFI; and the correlation between ADG and ROFC was r = -0.04 (sd = 0.72). The correlations between efficiency traits were weak or not correlated, in addition to presenting considerable standard deviations. This indicates that the FE-traits in calves were independent of the FE-traits in cows, and large sd indicates that larger sample size was needed to replicate the results. The genetic correlations can be seen in Figure 5 and their sd in Figure 6.

4.3.2.1. Genetic correlations between FE traits in calves and cows

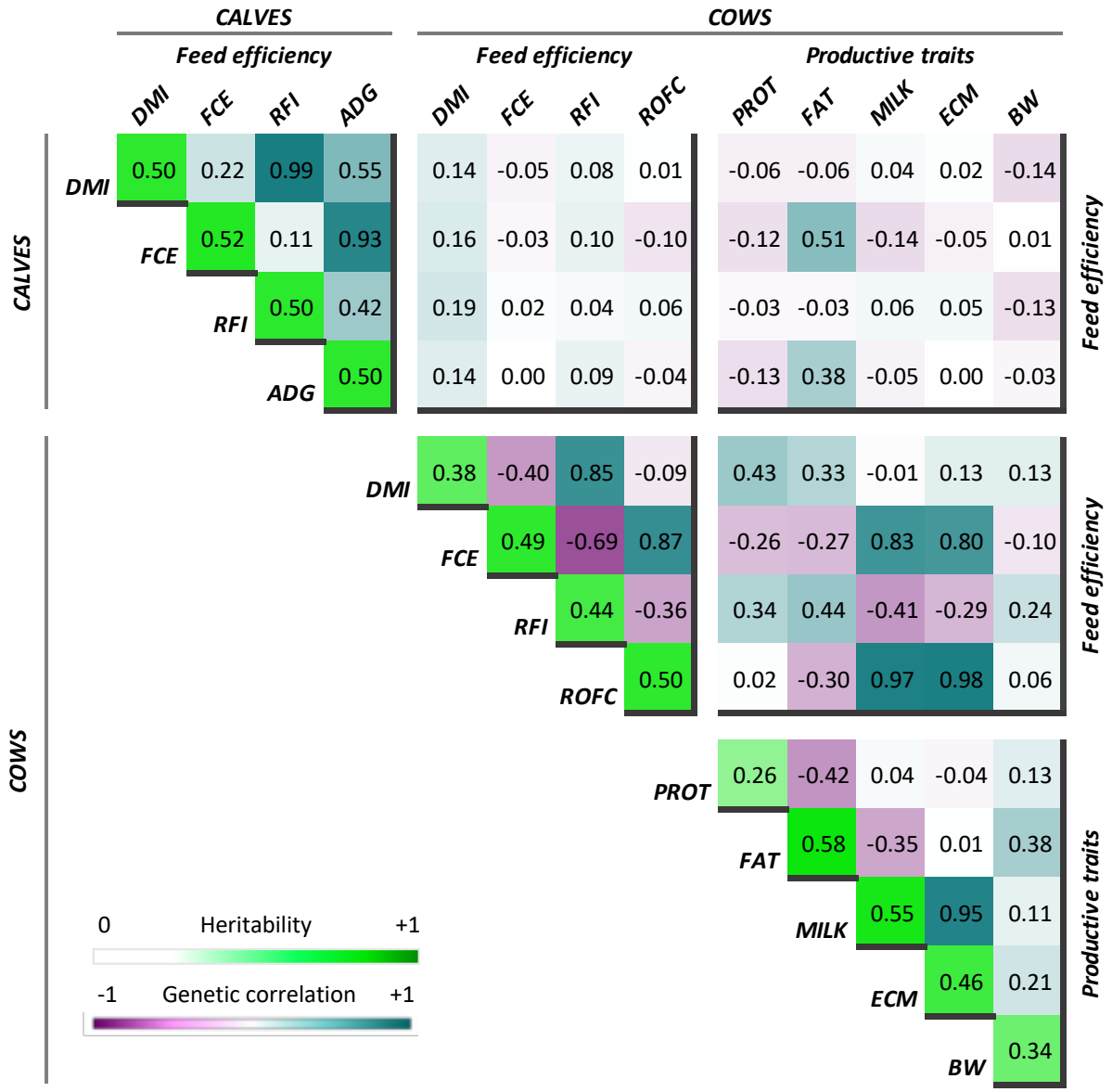


Figure 5. Heritabilities and genetic correlations for FE and productive traits in cows and calves. The green diagonal shows the heritability for each trait.

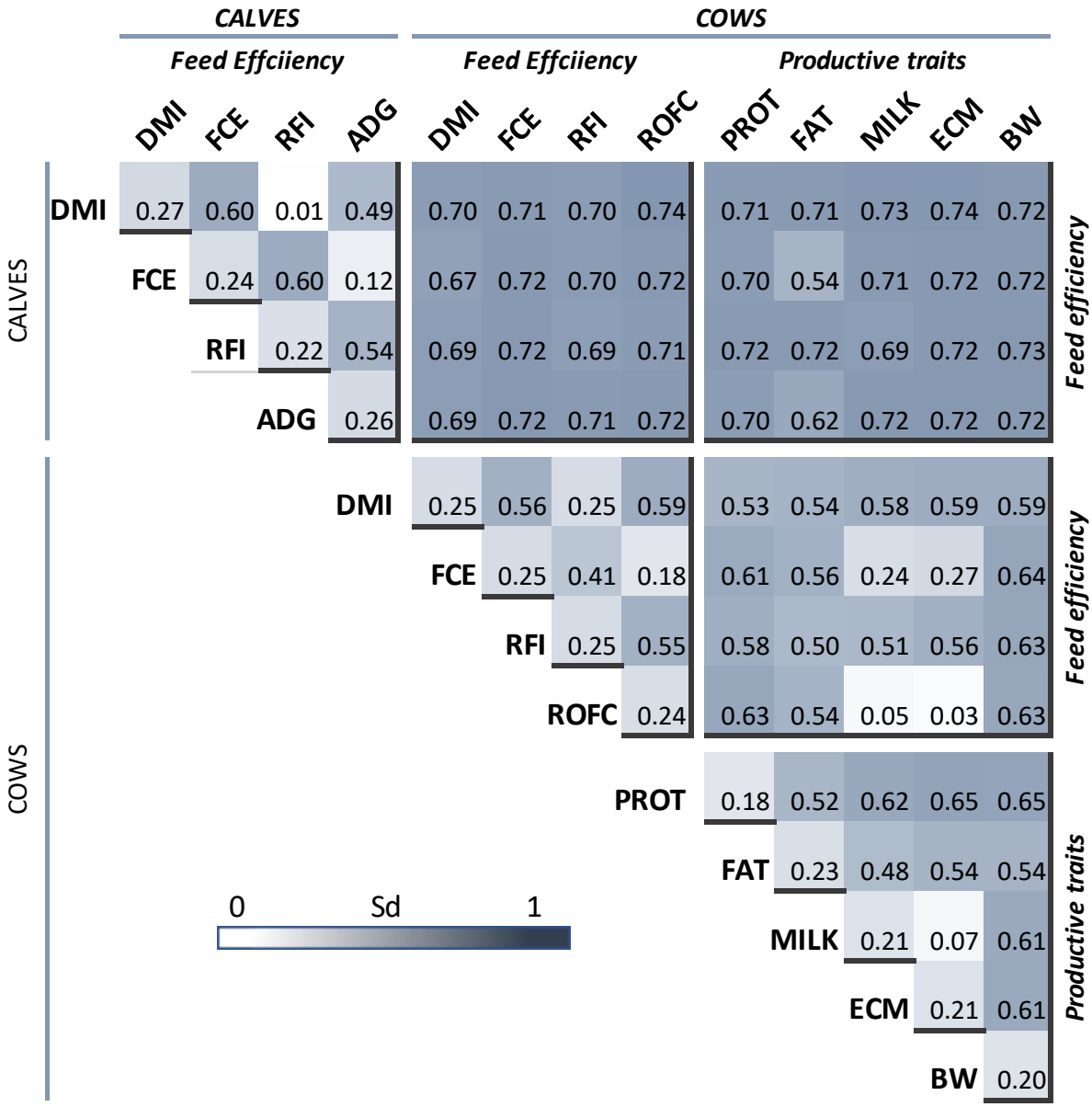


Figure 6. sd of the MCMC results for each heritability and genetic correlation.

4.4. Genome wide association study

4.4.1. Statistical power of the sample size

The data set used in this master thesis was too small to implement GWAS. Taking this into account, the objective of this section was to estimate the statistical power of the sample size to measure the reliability of the genome-wide association results (showed later). For this, the following parameters were considered as priors:

- The QTL allele frequencies were 0.5 for both p and q.
- Optimistically, we assume a linkage disequilibrium (D) of 0.25.

Then, three points were calculated: The Bayes factor (B) for the sample size (n) in this study, the power of the sample size (Pw) if B = 20 (that explain 5% and 1% of the variance), and n required if Pw = 0.80 and B = 20. The results are explained below and detailed in Table 9.

a) Bayes factor (B) equivalent to the sample size in this study (n).
   - In calves n = 30. In cows n = 70 (for FCE, RFI and ROFC) and n = 103 (for DMI). In both cases (calves and cows) B was less than 1 (Table 9), which suggests that the strength of the evidence was very small.

b) *Pw* that explains 5% and 1% of the variation (h_Q^2) assuming B = *20.*
   - For h_Q^2 = 0.05: Pw= 0.05, 0.09 and 0.14 for n = 30, 70 and 103 respectively.
   - For h_Q^2 = 0.01: Pw= 0.02, 0.01 and 0.01 for n = 30, 70 and 103, respectively

c) n required for a Pw = 0.8 and a B = 20.
   - when h_Q^2 = 0.05, n required was 415.
   - When h_Q^2 = 0.01 n required was 2566. Figure 7 shows an illustration of the relation between B and n for h_Q^2 = 0.05 and 0.01.

Table 9. Results of the analyses for the statistical power of the sample size.

Gr	n	p	q	D	h_Q^2	Trait	B	$Pw_{B_{20}}$	$n_{B_{20}}$
Calves	30	0.5	0.5	0.25	0.05	DMI, FCE, RFI, ADG	0.22	0.052	415
Calves	30	0.5	0.5	0.25	0.01	DMI, FCE, RFI, ADG	0.18	0.016	2566
Cows	70	0.5	0.5	0.25	0.05	FCE, RFI, ROFC	0.15	0.091	415
Cows	70	0.5	0.5	0.25	0.01	FCE, RFI, ROFC	0.08	0.012	2566
Cows	103	0.5	0.5	0.25	0.05	DMI	0.18	0.143	415
Cows	103	0.5	0.5	0.25	0.01	DMI	0.06	0.012	2566

Where GR was the group of study; n was the sample size in this study; p and q were the QTL allele frequencies; D was the linkage disequilibrium; h_Q^2 was the QTL heritability or the explained variation; TRAIT refers to the FE trait studied; $B_{P_{0.8}}$ was the Bayes factor for a statistical power of 0.8; $Pw_{B_{20}}$ was the statistical power of the sample size assuming a Bayes factor = 20; $n_{B_{20}}$ was the sample size required for Pw=0.8 and B=20. Note: DMI has n=103 because 33 cow's data were added for DMI genome-wide association.

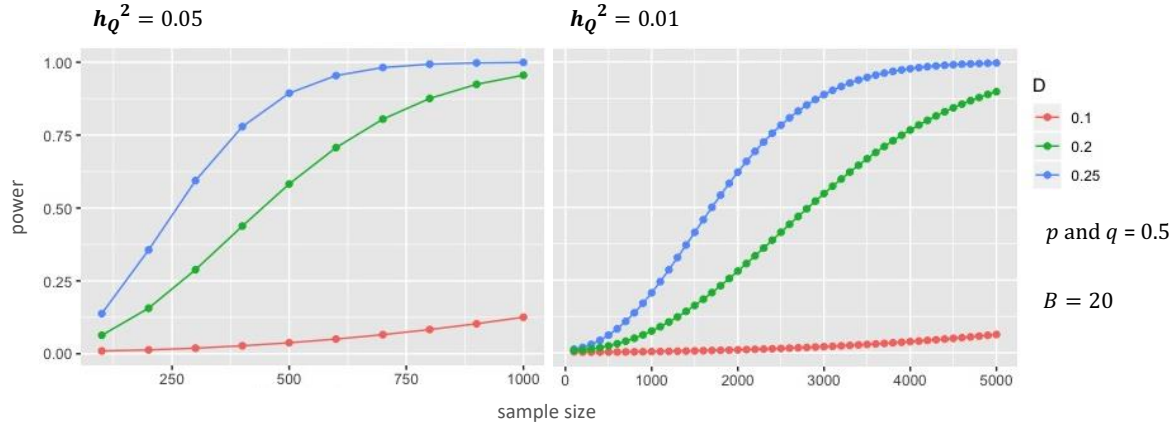


Figure 7. Power of the sample size explaining the 5% (left-plot) and 1% (right-plot) of the variation. Where D = linkage disequilibrium, this is shown by colour. This was calculated with an allelic frequency of 0.5 and the result is the equivalent to Bayes factor = 20.

These results suggest that the sample sizes were insufficient enough to give us reliable power. We have been optimistic with priors to estimate the strength of the evidence in the genome-wide association studies. However, the population size in this study did not represent a reliable power of the sample size. Therefore, genome association results should be interpreted with caution.

4.4.2. Models

4.4.2.1. Frequentist GWAS

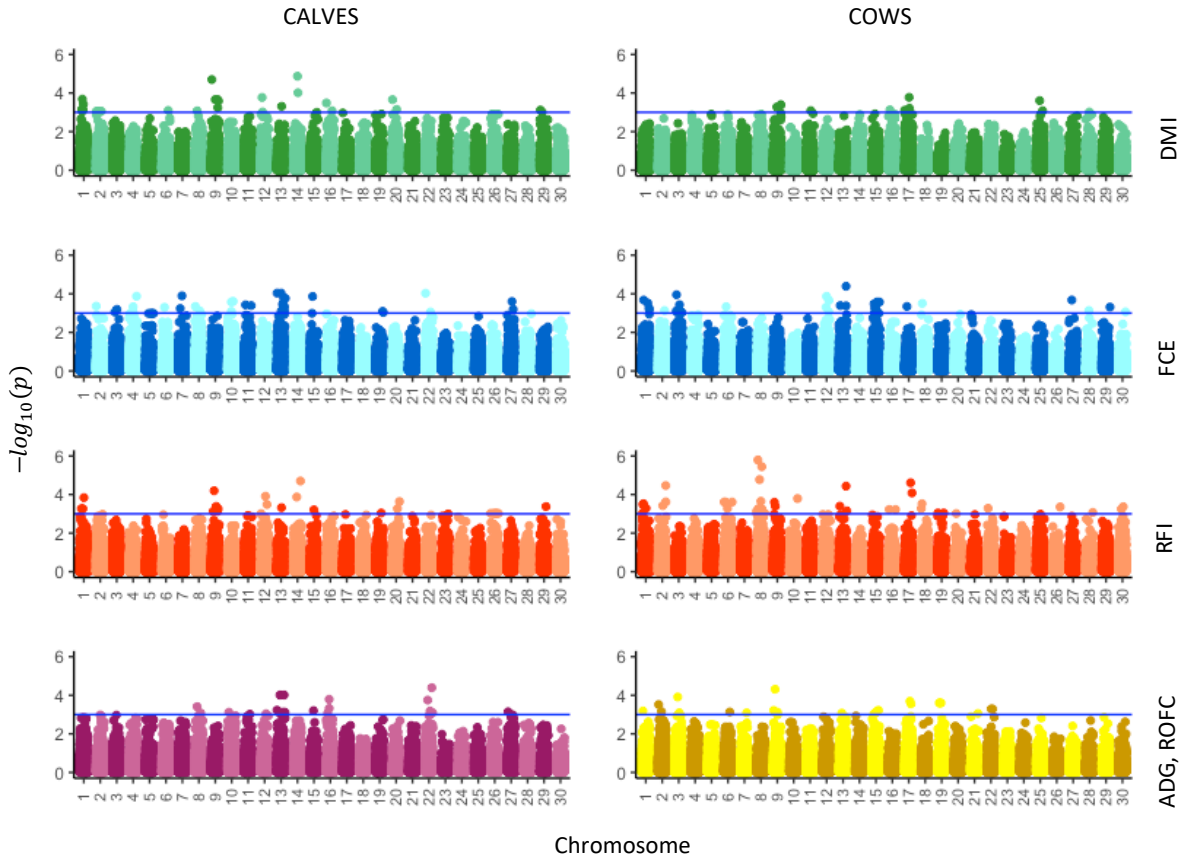


Figure 8. FE Manhattan plots from frequentist GWAS. Above the blue line, are the SNPs with P-value < 0.001

GWAS results were discussed through the SNP significance as follow:

Significant SNPs ($P < \alpha = 0.001$) for FE traits in GWAS

An alpha (α) 0.001 was established, being $p < 0.001$ a significant value for the SNP effect. Here, 28, 35, 25 and 22 significant SNPs were found for DMI, FCE, RFI, ADG, respectively in calves. Whereas 12, 34, 44 and 24 significant SNPs were found for DMI, FCE, RFI and ROFC, respectively in cows. For DMI in calves, 28 SNPs were identified on 13 different chromosomes; whereas in cows 12 SNPs identified were distributed on 6 different chromosomes. For FCE in calves, 35 SNPs were identified along 13 different chromosomes, whereas 34 SNPs were identified in cows, distributed on 12 different chromosomes. For RFI in calves, 25 the SNPs identified were distributed on 10 different chromosomes whereas in cows 44 SNPs were found on 17 different chromosomes. For ADG in calves, 22 SNPs were

found on 10 different chromosomes. For ROFC in cows, 24 significant SNPs were found on 12 different chromosomes. Unfortunately, no common SNP were found for any FE traits between cows and calves.

Table 10. Significant SNPs for P-values in GWAS

	Calves				Cows			
P-value	**DMI**	**FCE**	**RFI**	**ADG**	**DMI**	**FCE**	**RFI**	**ROFC**
0.001	28	**35**	25	22	12	34	**44**	24

Higher values are desired, the trait with the highest significant SNPs is marked in boldface

Number of SNPs in common between traits ($P < \alpha = 0.001$) in GWAS

In calves, DMI and RFI had more common SNPs between them. These sixteen SNPs (chromosome, and nearest gene in parenthesis) were *Hapmap43629-BTA-60810* (1), *BTB-01146938* (1), *ARS-BFGL-NGS-116361* (1), *ARS-BFGL-NGS-93995* (9), *ARS-BFGL-NGS-15511* (9), *ARS-BFGL-NGS-17690* (9), *Hapmap23835-BTA-161158* (9), *ARS-BFGL-NGS-113524* (9), *ARS-BFGL-NGS-110434* (12), *ARS-BFGL-NGS-60282* (12), *Hapmap60144-rs29013559* (13), *UA-IFASA-5750* (14), *Hapmap23726-BTC-051363* (14), *ARS-BFGL-NGS-44829* (20), *BTA-86837-no-rs* (20), *ARS-BFGL-NGS-64656 (29)*. Also, FCE had more common SNPs with and ADG, these fourteen SNPs were *ARS-BFGL-NGS-112477* (10), *Hapmap48260-BTA-24589* (11), *Hapmap47248-BTA-32461* (13), BTA-32556-no-rs (13), *ARS-BFGL-NGS-71025* (13), *BTA-115847-no-rs* (13), *Hapmap49962-BTA-32832* (13), *ARS-BFGL-NGS-2022* (13), *Hapmap49963-BTA-33040* (13), *ARS-BFGL-NGS-21830* (13), *Hapmap44175-BTA-98206* (15), *Hapmap52953-rs29025745* (22), *Hapmap35936-SCAFFOLD65654_2749* (22), *UA-IFASA-3305 (27)*. DMI and ADM had only one SNP in common, this was *Hapmap54267-rs29023167* (16). No more SNPs were found in common for other traits.

In cows, DMI and RFI had two common SNPs, these were *Hapmap48321-BTA-40830* (17) and *Hapmap49910-BTA-20754* (17); of them, none were found in common with the equivalent comparison in calves. FCE had seven SNPs in common, these were *Hapmap41492-BTA-26349* (3), *ARS-BFGL-NGS-118243* (3), *Hapmap50605-BTA-16738* (6), *ARS-BFGL-NGS-12483* (15), *ARS-BFGL-NGS-61425* (15), *ARS-BFGL-NGS-41288* (15), *ARS-BFGL-NGS-14291* (17); of them, none were found in common with the equivalent comparison in calves. No more SNPs were found in common for other traits.

In both cows and calves, DMI and RFI had more common SNPs, and FCE shared more SNPs with ADG (ROFC in cows). No SNPs have been repeated in the four FE-trait (DMI, FCE, RFI, ADG) in both calves and cows (Figure 9).

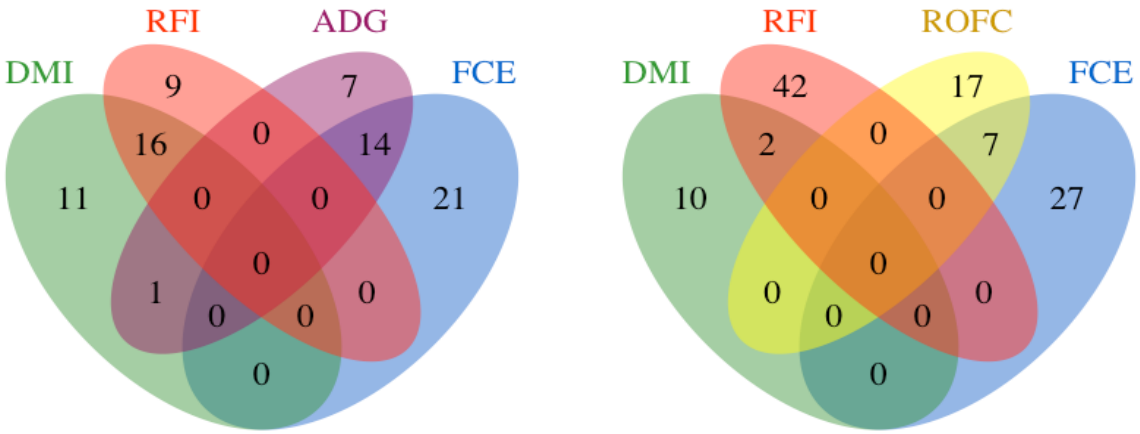


Figure 9. Venn Diagrams from common SNPs for FE in GWAS model. significant SNPs with alpha < 0.001 are shown. In the left, SNPs for calves, in the right, SNPs for cows.

The most significant SNPs and their closest genes and biological functions.

The most significative SNPs (the lowest *p*-value) were selected. Then, the closets genes and their associated biological function were found. The results are discussed below, and the SNPs, genes and biological function are shown in Table 11 and Table 12 for calves and cows respectively.

Most significant SNPs in calves in GWAS

The markers associated to FE in calves were found on chromosomes 9, 13, 14 and 22. All significant SNPs (3 SNPs) for FCE were also significant for ADG, and two of three SNPs significant for DMI were significant for RFI, this corroborates the high genetic correlations between these traits (Figure 5). Genes linked to DMI and RFI were involved in immune response, nervous development and protein synthesis; whereas the FCE and ADG genes were involved with enzymatic activity (aminopeptidase), molecular processes (transcription and transduction) and metal ion binding.

Cole et al. (2011) showed an association of milk and fat production with the *GNAS* gene; in this study the same gene was associated with FCE and ADG; this makes sense because FCE was strongly correlated to milk production and ADG (Figure 5). Besides, the genetic

correlation of FCE_{calves} and ADG_{calves} with FAT_{cows} was positive (Figure 5). El-Halawany et al. (2017) and Perez et al. (2018) have associated milk production and fertility with the *CSMD3* gene. In this study, the same genes were associated with DMI and RFI, this is interesting because these production traits should be related to FCE, although it was also due to DMI and FCE were correlated in calves. Mallikarjunappa et al. (2018) and de las Heras-Saldana et al. (2019) show *EIF3H* associated with immunity and RFI respectively. In this study, the same gene was associated with DMI and RFI, this suggests again that immunity plays a vital role in feed intake and metabolism.

Table 11. SNPs with lowest P-value and their nearest genes for FE traits in calves in GWAS.

					REFERENCE		
Ch	**Marker**	**Trait**	**QTL**	**Function**	**Author**	**Associated trait**	**Specie**
9	ARS-BFGL-NGS-113524	DMI RFI	SAMD3*	Immune response			
13	ARS-BFGL-NGS-21830	FCE ADG	NPEPL1*	Aminopeptidase activity			
	Hapmap49963-BTA-33040	FCE ADG	GNAS*	Molecular transducer	Cole et al. (2011)	Milk, fat	Holstein
			NELFCD*	Transcription regulation			
14	Hapmap23726-BTC-051363	DMI RFI	CSMD3*	Regulation of dendrite development	El-Halawany et al. (2017)	Milk	Buffalo
					Perez et al. (2018)	Scrotal circumference and pregnancy	Nellore
	UA-IFASA-5750	DMI RFI	EIF3H**	Protein biosynthesis	Mallikarjunappa et al. (2018)	Antibody to Johne's disease	Holstein
					De Las Heras-Saldana et al. (2019)	RFI	Angus
22	Hapmap52953-rs29025745	FCE ADG	FGD5*	Metal ion binding			

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

Most significant SNPs in cows in GWAS

In cows, the most significant SNPs were found on chromosomes 13, 17, 2, 8 and 9, and the related genes can be seen in Table 12. Hapmap48321-BTA-40830 and Hapmap49910-BTA-20754 SNPs match between DMI and RFI, whereas FCE and ROFC do not share genes within the most significant genes. The genes associated with RFI were involved in various

biological functions, including Insulin regulation, apoptosis processes, glycoprotein synthesis, cell development, spermatogenesis, organ development and immunity. The genes associated with FCE and ROFC were involved in the transcription process and cellular attachment, assembly and transport process.

PRUNE2 and *IL2* were associated with RFI; this association has also been reported by Lima et al. (2016) and Hou et al. (2012). *EPB41L1* and *SPRY1* genes that in this study were associated with RFI, in the investigations of Mudadu et al. (2016) and Zhou et al. (2019) have been associated with backfat thickness and gestation length; This could be due to feeding and saving intake could be affected by factors related to fat metabolism and energy expenditure by pregnancy. Three studies have referenced the *NCOA6* gene and two the *TP53INP2* gene, they related these to scrotal circumference in males, body fat, and milk fatty acids in females (Table 12); in this study, both genes have been related to FCE; this suggests that feed efficiency in cows could be influenced by genes involved in the production, lipid metabolism and fertility.

Among the QTLs reviewed in Table 11 and Table 12 no common genes (between most significant) were found between cows and calves.

Table 12. SNPs with lowest P-value and their nearest genes for FE traits in cows in GWAS.

					REFERENCE		
Ch	**Marker**	**Trait**	**QTL**	**Function**	**Author**	**Associated trait**	**Specie**
2	BTB-00093493	RFI	GRB14*	Insulin receptor regulation			
8	ARS-BFGL-NGS-17993	RFI	PRUNE2*	Apoptotic process	Lima et al. (2016)	RFI	Nellore
			GCNT1*	Glycoprotein biosynthetic			
	BTB-00631715	RFI	FOXB2*	Anatomical structure			
9	BTB-00389188	ROFC	TRAPPC3L*	Golgi vesicle-mediated transport			
			CALHM6*	cation channel activity			
			CALHM5*	cation channel activity			
13	ARS-BFGL-NGS-63663	RFI	CNBD2*	spermatogenesis			
			EPB41L1*	actomyosin and actin organization	Mudadu et al. (2016)	back fat thickness	Nellore
	Hapmap54034-rs29026486	FCE	PIGU*	Attachment of GPI anchor to protein			
			NCOA6*	Transcription coactivator activity	Irano et al. (2016)	Scrotal circumference	Nellore
					Júnior et al. (2016)	Back fat thickness	Nellore
					Olsen et al. (2017)	milk fatty acids	Dairy cattle
					Knutsen et al. (2018)	milk fatty acids	Dairy cattle
			TP53INP2*	Autophagosome assembly	1. Irano et al. (2016)	1. scrotal circumference	Nellore
17	Hapmap48321-BTA-40830	DMI RFI	SPRY1**	animal organ development	Zhou et al. (2019)	1. gestation length	Xinjiang Brown cattle
	Hapmap49910-BTA-20754	DMI RFI	IL21*	adaptive immune response	1. Hou et al. (2012)	1. RFI	1.Holstein cattle
					Gurgul et al. (2019)	Inmune response	Polish cattle

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

4.4.2.2. Bayes LASSO association

For the Bayes LASSO analysis, the comparison was made through the SNP effect and the discussion of the results are shown below.

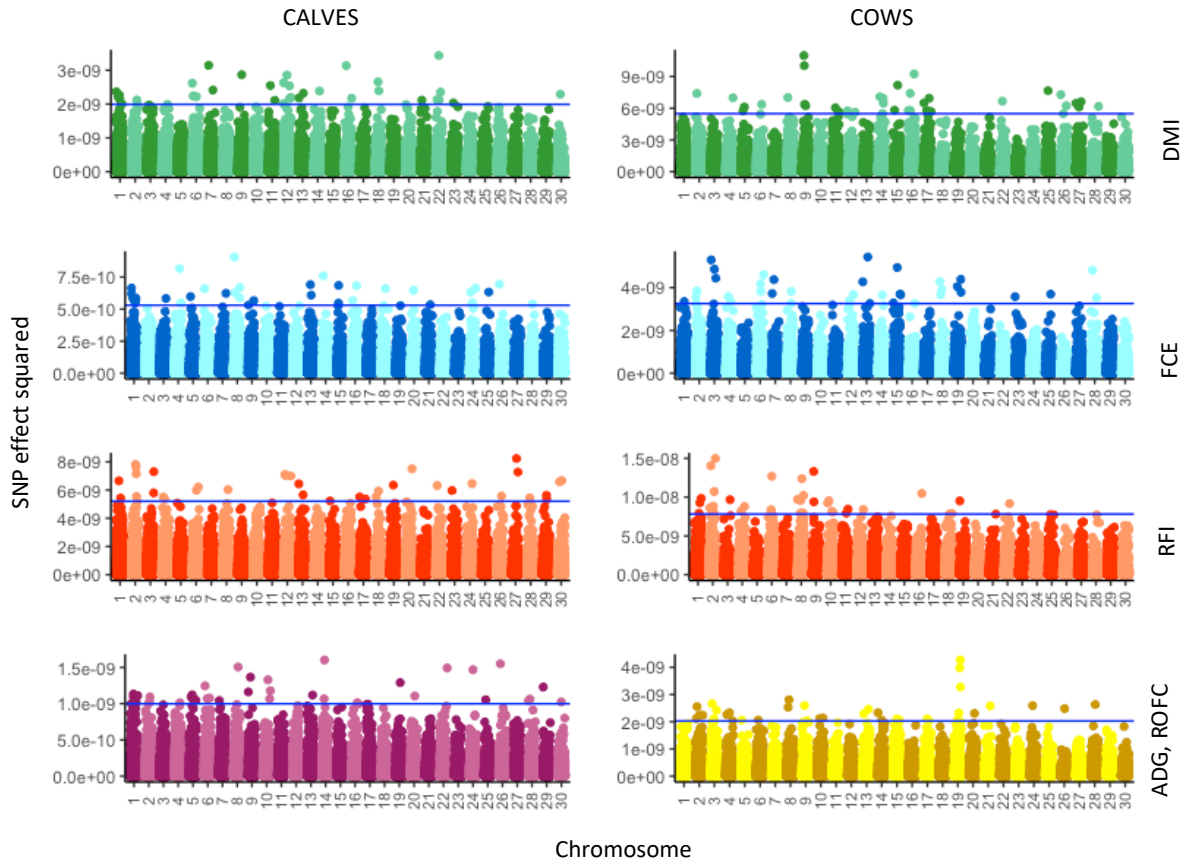


Figure 10. FE Manhattan plots from Bayesian-LASSO association. Above the blue line, are the SNPs with quantile > 0.999.

Significant SNPs (quantile> 0.999) for FE traits in Bayes LASSO

The analyses of this section were performed based on the SNP effect. The additive genetic variance of the SNPs was estimated being always <1% (Figure 12). Thus, these results were not used in the comparative analyses. Approximately, 35 SNPs with larger effect (quantile> 0.999) from each trait were selected, these are shown in Figure 10. For DMI, the SNPs identified in calves were distributed on 13 different chromosomes; whereas in cows the SNPs were distributed on 17 different chromosomes. Unlike GWAS and random forest, in Bayes LASSO 1 SNP in common between calves and cows was found, this SNP was *Hapmap48117-BTA-90454* and it was found on chromosome 12. For FCE, the SNPs identified in calves were distributed on 20 different chromosomes, whereas in cows they were distributed on 17 different chromosomes and do not share SNPs between cows and calves. For RFI in calves, the SNPs identified were distributed on 21 different chromosomes whereas in cows they were found on 15 different chromosomes, one of them also was found in common between cows and calves, this SNP was *BTA-43831-no-rs* and it was found on chromosome 18 (none SNP in common was found between cows and calves in GWAS and RF). For ADG in calves,

the SNPs were found on 22 different chromosomes. Finally, for cow ROFC, significant SNPs were found on 18 different chromosomes. For the interpretation of these results, one aspect to consider is that lambda convergence was not too good in both calves and cows (Annex 9 and Annex 10).

Number of SNPs in common between traits (quantile > 0.999) in Bayes LASSO

In calves, Only DMI and RFI shared two SNPs in common, these SNPs were *ARS-BFGL-NGS-41287* (1) and *BTA-86837-no-rs* (20). No more SNPs were found in common for other traits.

In cows, DMI had three SNPs in common with FCE, these were *ARS-BFGL-NGS-41523* (2), *ARS-BFGL-NGS-103734* (17), and *ARS-BFGL-NGS-111019* (28). DMI and RFI had six common SNPs, these were *BTB-00901654* (8), *ARS-BFGL-NGS-85644* (8), *BTB-01286081* (16), *BTB-00393938* (16), *Hapmap26379-BTA-130999* (17), *ARS-BFGL-NGS-75936* (18). DMI had two SNPs in common with ROFC, these were *ARS-BFGL-NGS-54368* (16) and *ARS-BFGL-NGS-111019* (28). FCE had two SNPs in common with RFI, there were *ARS-BFGL-NGS-95329* (2) and BTB-01195369 (3). FCE had six SNPs in common with ROFC, these were *BTB-01450068* (4), *BTB-00493207* (12), *ARS-BFGL-NGS-61425* (15), *ARS-BFGL-NGS-82204* (19), *ARS-BFGL-NGS-118018* (19), *ARS-BFGL-NGS-111019* (28). RFI had one SNP in common with ROFC, this was *Hapmap34677-BES4_Contig489_1116* (8). DMI, FCE y ROFC had one SNP in common, this was *ARS-BFGL-NGS-111019* (28). No more SNPs were found in common for other traits and no SNPs were found in common with the equivalent comparisons in calves.

In both cows and calves, DMI and RFI had more common SNPs, while FCE shared more SNPs with ROFC only in cows, this differs from the analyses found in GWAS. No SNPs have been repeated in the four FE-trait (DMI, FCE, RFI, ADG) in both calves and cows (Figure 11).

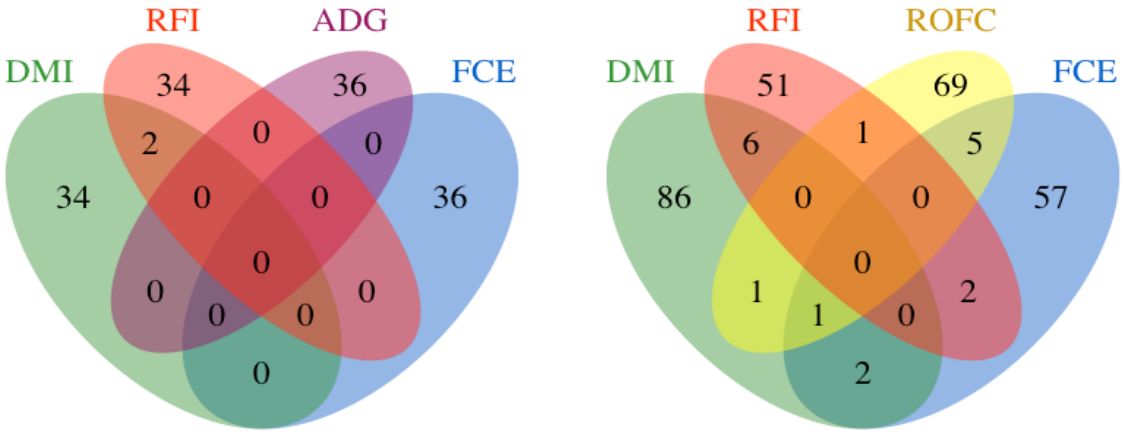


Figure 11. Venn Diagrams of common SNPs for FE in Bayes LASSO association. SNPs with highest effect in the quantile > 0.999 are shown, In the left, SNPs for calves, in the right, SNPs for cows.

Most significant SNPs in calves (quantile > 0.9999) in Bayes LASSO

In calves, within the SNPs with the most significant effect (q> 0.9999) there were no SNPs in common between FE traits, this differs with the traditional GWAS analysis. The biological functions in which the genes were involved were diverse. RFI and DMI showed association with the immune response, behaviour and bone development; whereas the genes most associated with FCE and ADG were involved in cell replication, transport and calcium regulation. Each gene and their most associated functions are detailed in Table 13.

Only one gene associated with RFI was referenced in another study, this gene was INSIG2 associated with MILK in Deng et al. (2016). The PCDH7 gene was associated with DMI, De Lima et al. (2017) found a relationship between this gene and RFI (this gene had significant activity in the liver). Five other genes that were associated with DMI in this analysis, in other studies these were associated with body size, body fat, abomasum displacement, meat quality and bone development. Association of FCE with the ARHGAP20 and DOCK1 genes was found, the studies of Zhang et al. (2017) and Neupane et al. (2017) associated these genes with body size and fertility in heifers respectively. The ATP2B2 gene that in this study was associated with ADG, Gonçalves et al. (2018) associates it with beef tenderness in Nellore cattle. In general, several significant genes that were appreciated in this analysis, in other studies have been associated with production and fertility. The genes found in this study were not matched with genes in GWAS analyses; this shows the role and importance of several genes within different physiological routes influencing the FE of animals.

Table 13. SNPs with the highest effect (quantile > 0.9999) and their nearest genes for FE traits in calves in Bayes LASSO.

					REFERENCE		
Ch	Marker	Trait	QTL	Associated Function	Author	Associated trait	Specie
2	ARS-BFGL-NGS-20993	RFI	INSIG2**	Cholesterol metabolic process	Deng et al. (2016)	MILK	Chinese buffaloes
	ARS-BFGL-NGS-88046	RFI	EN1*	Behaviour and brain development			
			MARCO*	Scavenger receptor activity			
3	Hapmap50068-BTA-69023	RFI	BEND5**	Transcription, DNA-templated			
6	BTA-76341-no-rs	DMI	PCDH7*	Cell adhesion	An et al. (2020)	Body size	Simmental cattle
					De Lima et al. (2017)	RFI	Nellore cattle
7	Hapmap53565-rs29013278	DMI	CETN3**	Calcium ion binding	1. Hardie et al. (2017)	Metabolic body weight	Holstein cattle
8	Hapmap49424-BTA-105436	ADG	ROR2	ATP binding			
9	Hapmap46780-BTA-18414	DMI	HTR1B**	Bone remodeling	Duncombe (2016)	Carcass fat	Beef cattle
12	BTB-02009715	DMI	SLITRK5*	Adult behavior	1. Biffani et al. (2014)	abomasum displacement	Holstein cattle
13	ARS-BFGL-NGS-23363	FCE	SLC24A3	Calcium ion transport			
15	Hapmap41192-BTA-16797	FCE	ARHGAP20*	Signal transduction	Zhang et al. (2017)	Body size	Chinese Holstein
18	ARS-BFGL-NGS-33562	DMI	WWOX*	Cellular response to transforming growth factor beta stimulus	Lee et al. (2018)	meat colour bone	Korean cattle
					Ramayo-Caldas et al., (2016)	metabolism	French cattle
22	ARS-BFGL-NGS-110768	DMI	SLC6A11	Neurotransmitter uptake			
	ARS-BFGL-NGS-62254	ADG	ATP2B2	Cellular calcium ion homeostasis	Gonçalves et al. (2018)	Beef Tenderness	Nellore cattle
			SEC13*	COPII-coated vesicle budding			
24	BovineHD4000000094	ADG	PIK3C3**	Autophagosome assembly			
26	ARS-BFGL-NGS-74523	FCE	C26H10orf90**	Regulation of centriole replication	Neupane et al. (2017)	heifer fertility	Beef cattle
			DOCK1*	Cell migration			
	Hapmap58513-rs29024371	ADG	VAX1*	Brain development	Cai et al. (2020)	Milk and fat yield	Holstein cattle
			KCNK18*	Ion transport			
27	ARS-BFGL-NGS-114154	RFI	PPP1R3B*	Carbohydrate metabolism			
	ARS-BFGL-NGS-45124	RFI	MCPH1**	Cerebral cortex development			

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

Most significant SNPs in cows (quantile > 0.9999) in Bayes LASSO

In cows, the most significant SNPs were found on various chromosomes (between 2 and 28). The functions of the nearest genes were involved with various molecular and biological routes and processes, and there was no clear pattern that relates these with a specific FE trait. However, these genes can be classified as follow: DMI associated genes were involved in system nervous, transcription, fertility and cellular maintenance process; FCE associated genes were involved in regulatory functions, oxidoreduction and immunity; RFI associated genes were involved in ionic transport and various cellular functions, and ROFC-associated genes were involved in the development of the central nervous system, protein localization and activation. Between the most significant SNPs, no marker shared between FE traits were found.

From the SNPs whit the highest effect, seven referenced genes have been found in other studies. However, none have been directly associated with FE traits. The NCKAP5, TBCK and NPNT genes, which in this study were found to be significant for RFI, the first has been referenced by Wu et al. (2016) in association with feet and legs disorders in multi-breed cattle, and the other two have been referenced by Carvalho et al. (2020) in association with muscle tissues and cartilage development in Nellore cattle. The USP24, NCOA6, TP53INP2 and PSAP genes in this study were associated with FCE; Cai et al. (2020), Olsen et al. (2017), Irano et al. (2016) and Guo et al. (2016) associated these genes with milk yield (dairy cattle), milk fatty acids (dairy cattle), scrotal circumference (Nellore cattle) and meat quality (Simmental cattle), respectively. The CDH2 gene was significant for ROFC and Zhou et al. (2019) relates it to milk production in dual-purpose cattle. Traditional GWAS with Bayes LASSO models match one SNP, this was Hapmap54034-rs29026486 and it was significant for FCE in cows.

Table 14. SNPs with the highest effect (quantile > 0.9999) and their nearest genes for FE traits in cows in Bayes LASSO.

					REFERENCE		
Ch	**Marker**	**Trait**	**QTL**	**Associated Function**	**Author**	**Associated trait**	**Specie**
2	ARS-BFGL-NGS-115279	RFI	NCKAP5**	microtubule formation	Wu et al. (2016)	feet and legs disorders	Multi-breed cattle
	ARS-BFGL-NGS-95329	RFI	DPP10**	ion channel binding			
3	BTB-01195369	FCE	USP24**	protein deubiquitination	1. Cai et al. (2020)	Milk yield	Holstein cattle
	Hapmap53424-rs29019267	FCE	CYB561D1*	oxidoreductase activity			
6	ARS-BFGL-NGS-118535	RFI	TBCK* NPNT*	intracellular protein transport calcium ion binding	Carvalho et al. (2020)	muscle tissues and cartilage development	Nellore Cattle
9	ARS-BFGL-NGS-29072	DMI	MCHR2*	neuropeptide signaling			
	BTB-00393938	DMI	FHL5*	transcription by RNA polymerase II			
	BTB-01286081	RFI	GPR63*	G protein-coupled receptor signaling			
13	Hapmap54034-rs29026486	FCE	NCOA6*	Transcription coactivator activity	Irano et al. (2016)	scrotal circumference	Nellore cattle
					Júnior et al. (2016)	back fat thickness	Nellore cattle
					Olsen et al. (2017)	Milk fatty acids	Dairy cattle
					Knutsen et al. (2018)	milk fatty acids	Dairy cattle
			TP53INP2*	Autophagosome assembly	Irano et al. (2016)	scrotal circumference	Nellore cattle
15	ARS-BFGL-NGS-61425	FCE	SERGEF	regulation of protein secretion			
16	ARS-BFGL-NGS-15747	DMI	NPHP4**	flagellated sperm motility			
19	ARS-BFGL-NGS-109844	ROFC	ALDH3A2*	central nervous system development			
	ARS-BFGL-NGS-82204	ROFC	AKAP10*	protein localization			
			ULK2*	activation of protein kinase activity			
24	BovineHD2400007677	ROFC	CDH2**	Calcium-dependent cell adhesion protein	1. Zhou et al. (2019)	Fat yield	Xinjiang Brown cattle
25	ARS-BFGL-NGS-7049	DMI	KCTD13*	cell migration			
28	ARS-BFGL-NGS-111019	FCE	PSAP*	antigen processing and presentation	Guo et al. (2016)	carcass composition and meat quality	Chinese Simmental-cross cattle

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

Proportion of genetic additive variance in Bayes LASSO

The proportion of the genetic variance explained by the SNPs was estimated. Their results presented very low values:

- The highest value for DMI was 1.69e-09 in calves and 4.98e-09 in cows.
- The highest value for FCE was 4.19e-10 in calves and 2.71e-09 in cows.
- The highest value for RFI was 3.89e-09 in calves and 7.41e-09 in cows.
- The highest value for ADG in calves was 7.86e-10 and the highest value for ROFC in cows was 2.13e-09.

The genetic variance explained can be seen in Figure 12.

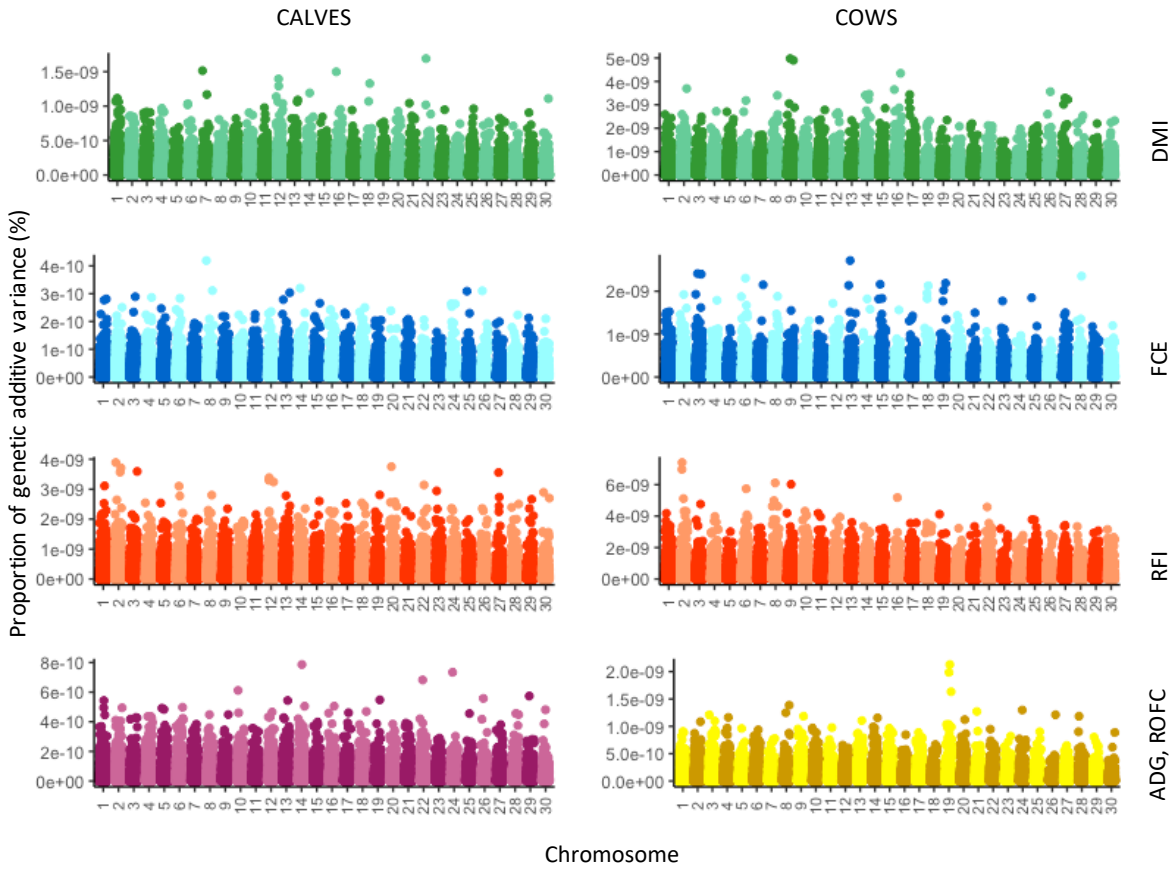


Figure 12. FE Manhattan plots of proportion of genetic additive variance from Bayesian-LASSO.

4.4.2.3. Random forest

For the random forest analysis, the comparison was made through the SNP importance and the results are shown and discussed below.

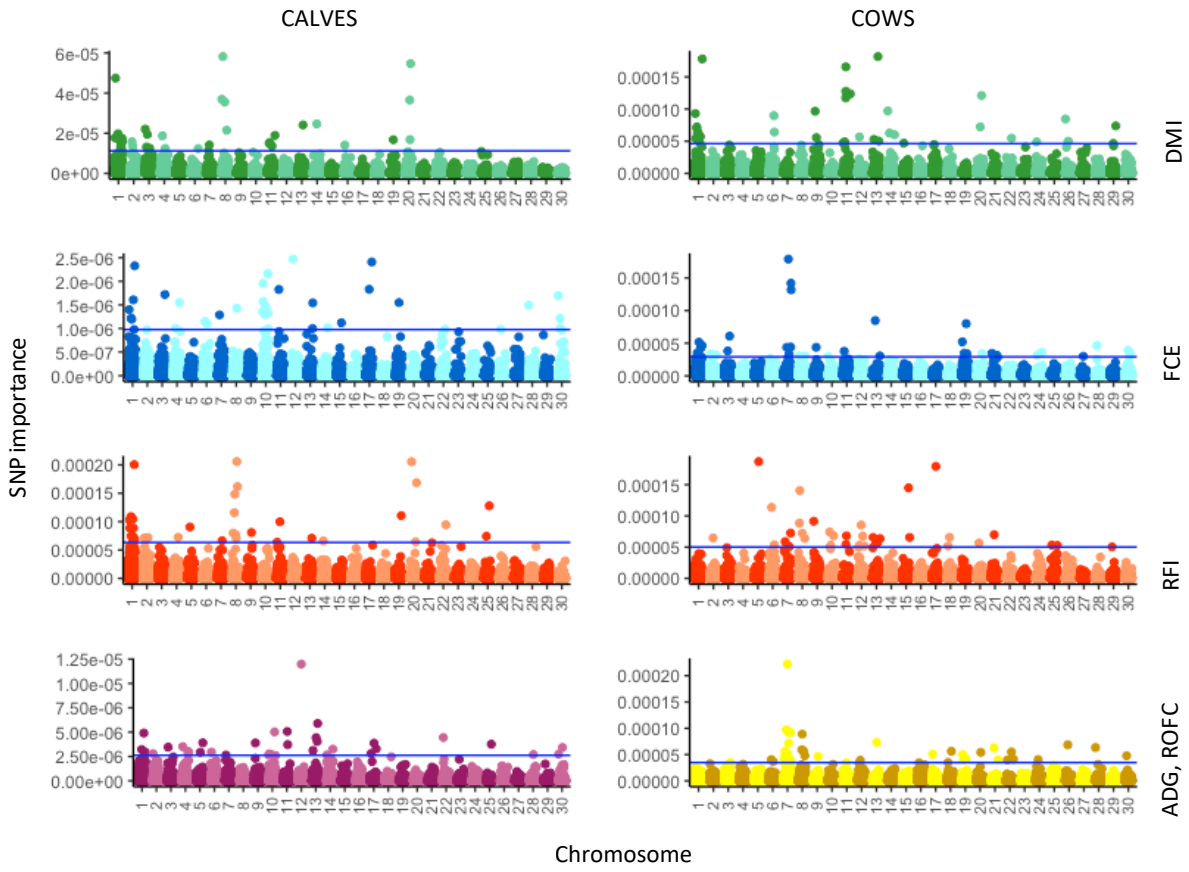


Figure 13. FE Manhattan plots from random forest association. Above the blue line are the SNPs with quantile > 0.999.

Significant SNPs (quantile> 0.999) for FE traits in random forest

Approximately the 35 more important SNPs from each trait were selected (Figure 13). For DMI, the SNPs identified in calves were dispersed on 13 different chromosomes; in cows, the SNPs were also distributed on 13 chromosomes but different from those on calves. For FCE, the SNPs identified in calves were founded on 19 different chromosomes, whereas in cows they were distributed on 15 different chromosomes. For RFI in calves, the SNPs identified were distributed on 16 different chromosomes; whereas in cows they were found on 16 different chromosomes. For ADG in calves, the SNPs were found on 18 different chromosomes. Finally, for ROFC in cows, the most important SNPs were found on 16 different chromosomes.

Number of SNPs in common between traits (quantile > 0.999) in random forest

In calves, DMI and RFI had twenty-two SNPs in common, these were ARS-BFGL-NGS-98459 (1), BTA-114651-no-rs (1), BTB-00032205 (1), Hapmap41227-BTA-32644 (1), Hapmap39468-BTA-120746 (1), BTB-00031796 (1), ARS-BFGL-NGS-26880 (1), Hapmap34848-BES1_Contig523_1341 (2), BTB-01330347 (4), ARS-BFGL-NGS-94147 (7), ARS-BFGL-NGS-57673 (8), ARS-BFGL-NGS-105601 (8), Hapmap45972-BTA-102617 (8), BTB-01364009 (8), BTA-119672-no-rs (11), Hapmap41707-BTA-99303 (13), ARS-BFGL-NGS-100055 (14), ARS-BFGL-NGS-36291 (19), ARS-BFGL-NGS-20300 (20), ARS-BFGL-NGS-17910 (20), Hapmap60719-rs29027054 (20), and Hapmap24609-BTC-015462 (25). DMI and ADG had one SNP in common, this was BTB-01630036 (11). FCE and ADG had sixteen SNPs in common, these were ARS-BFGL-BAC-31482 (1), ARS-BFGL-NGS-118362 (2), Hapmap25108-BTA-18447 (4), ARS-BFGL-NGS-70470 (6), ARS-BFGL-BAC-12872 (10), ARS-BFGL-NGS-103742 (10), ARS-BFGL-NGS-31962 (10), ARS-BFGL-NGS-79766 (10), Hapmap44164-BTA-92933 (11), ARS-BFGL-BAC-15043 (12), Hapmap52923-rs29015102 (13), ARS-BFGL-NGS-113153 (17), ARS-BFGL-NGS-4366 (17), ARS-BFGL-NGS-110727 (28), Hapmap60788-rs29017234 (X), Hapmap59288-rs29021774 (X). No more SNPs were found in common for other traits.

In cows, FCE and RFI had two SNPs in common, these were *BTB-00283498* (6) and BTB-00770436 (20). FCE and ROFC had twelve SNPs in common, these were BTB-01558306 (7), ARS-BFGL-NGS-108870 (7), Hapmap47162-BTA-103817 (7), ARS-BFGL-NGS-201 (7), BTB-01182727 (9), ARS-BFGL-NGS-113647 (13), ARS-BFGL-NGS-118018 (19), ARS-BFGL-NGS-109844 (19), BTB-00770436 (20), UA-IFASA-6258 (21), Hapmap47360-BTA-63966 (28), Hapmap60265-rs29024291 (32). FCE, RFI and ROFC had one SNP in common, this was BTB-00770436 (20). No more SNPs were found in common for other traits and no SNPs were found in common with the equivalent comparisons in calves.

In calves DMI and RFI had more common SNPs, and FCE had more SNPs in common with ADG, whereas in cows only FCE with ROFC share more common SNPs. (Figure 14).

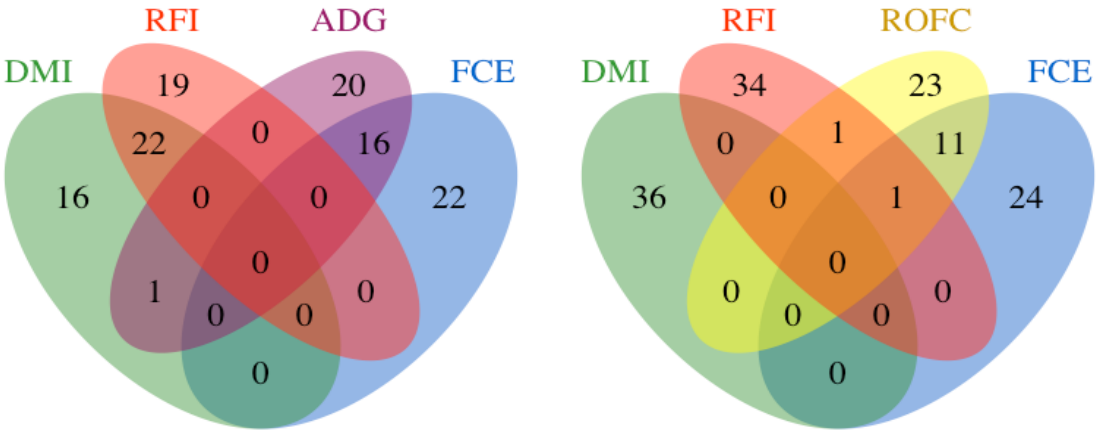


Figure 14. Venn Diagrams for common SNPs for FE in random forest association. SNPs with highest effect in the quantile > 0.999 are shown, In the left, SNPs for calves, in the right, SNPs for cows.

Most significant SNPs in calves (quantile > 0.9999) in random forest

The closest genes to the most important SNPs for FE traits were selected and they were detailed in Table 15, these were distributed between chromosomes 1 and 20. DMI and RFI share three genes in common: SYK, HGE1 and DEPDC1B; of which, the first has been referenced by Buitenhuis et al. (2014) and associates it with milk fat. FCE and ADG share two genes in common: MDGA2 and MBNL2, and curiously, Yao et al. (2013) associate the first one with RFI in Holstein cattle, this shows the similarity between efficiency traits in calves even between FCE and RFI. The CAVIN4 gene had a significant effect on DMI, Uemoto et al. (2020) also associate this gene with DMI and CH4 per DMI units (CH4 / DMI) in Japanese Black steers. The LRRTM4 and GNAS genes were associated to ADG and these genes were associated with milk and fat production by Li et al. (2014) and Cole et al. (2011), respectively. Cai et al. (2020) also associate the LRRTM4 gene with milk yield and mastitis resistance. The GNAS gene was also significant in the analysis of traditional GWAS in calves.

In general, the genes associated with DMI and RFI were involved in fat cell differentiation, immune response, organ development, and cell maintenance functions. Genes related to FCE and ADG were involved in nervous system development, ion binding and transduction functions. So, as in previous analyses, FCE and ADG were referenced with association studies on productive traits, and one observation was that some RFI and DMI markers were also associated with productive traits.

Table 15. The most important SNPs and their nearest genes (quantile > 0.9999) for FE traits in calves in random forest association.

					REFERENCE		
Ch	Marker	Trait	QTL	Associated function	Author	Associated trait	Specie
1	BTB-00030894	FCE	GAP43	Nervous system development			
	BTB-00031796	DMI RFI	OSBPL11	Fat cell differentiation			
8	ARS-BFGL-NGS-105601	RFI	SYK	Immune response	Buitenhuis et al. (2014)	Milk fat	Danish Holstein
	Hapmap45972-BTA-102617	DMI	SYK**	Immune response	Buitenhuis et al. (2014)	Milk fat	Danish Holstein
			DIRAS2*	Signal transduction			
	BTB-01364009	DMI	CAVIN4*	Muscle organ development	Uemoto et al. (2020)	CH4/DMI and DMI	Japanese steers
10	ARS-BFGL-NGS-103742	FCE ADG	MDGA2	NA	Yao et al. (2013)	RFI	Holstein cattle
11	Hapmap44164-BTA-92933	ADG	LRRTM4*	Neurexins and neuroligins Relation with	Li et al. (2017)	Yearling weight	Korean cattle
					Cai et al. (2020)	Milk yield and mastitis resistance	Holstein cattle
					Li et al. (2014)	Saturated fatty acid	Chinese Holstein
12	ARS-BFGL-BAC-15043	FCE ADG	MBNL2*	Metal ion binding			
13	Hapmap49963-BTA-33040	ADG	GNAS*	Molecular transducer	Cole et al. (2011)	Milk and fat	Holstein cattle
17	ARS-BFGL-NGS-113153	FCE	NOC4L	Protein glycosylation			
20	ARS-BFGL-NGS-17910	RFI	NDUFAF2 *	Cellular respiration			
	Hapmap60719-rs29027054	DMI RFI	DEPDC1B *	Cell migration			

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

Most significant SNPs in cows (quantile > 0.9999) in random forest

In cows, the closest genes to the most important markers on FE traits were identified and they were detailed in Table 16. DMI and RFI did not match SNPs in this category and their associated genes were involved in behaviour, ion binding, transcription, and immune response functions. Particularly the IL10RA and IL21 genes, which have been referenced by Gurgul et al. (2019) and Salleh et al. (2018) who related these genes with RFI and the immune system. FCE shares three genes in common with ROFC, these were TSPAN17, EIF4E1B and TSPAN17, and they were associated with protein maturation, transduction and cell maintenance. Two common SNPs were found between traditional GWAS and random forest

association in cows; these markers were BTB-00631715 and Hapmap49910-BTA-20754, and their nearby genes were FOXB2 and IL21, respectively. In Table 16, some genes associated with DMI were compared to studies that associate them to fatty acids in milk, weaning weight, weight carcasses and feed intake.

Table 16. The most important SNPs and their nearest genes (quantile > 0.9999) for FE traits in cows in random forest association.

					REFERENCE		
Ch	**Marker**	**Trait**	**QTL**	**Associated function**	**Author**	**Associated trait**	**Specie**
1	BTB-00037209	DMI	EHHADH*	Fatty acid beta-oxidation	Shi et al. (2019)	milk fatty acids	Chinese Holstein
			C1H3orf70*	Circadian behaviour			
5	ARS-BFGL-NGS-3503	RFI	ATXN7L3B**	Regulation of gene expression	Li et al. (2017)	weaning weight	Korean cattle
			CAPS2**	Calcium ion binding			
7	ARS-BFGL-NGS-108870	FCE ROFC	TSPAN17*	Protein maturation			
	ARS-BFGL-NGS-201	ROFC	STK32A	Intracellular transduction			
	BTB-01558306	FCE ROFC	EIF4E1B	RNA 7-methylguanosine cap binding			
8	BTB-00631715	RFI	FOXB2*	Anatomical structure			
11	BTB-01763350	DMI	CNOT11*	Deadenylation of mRNA			
13	ARS-BFGL-NGS-113647	FCE	NFATC2*	DNA binding			
	BTB-01329459	DMI	ZNF217*	Regulation of transcription	Mullen et al. (2011)	increased cow carcass weight	Holstein cattle
			TSHZ2*	Transcription by RNA polymerase II	Gan et al. (2019)	triiodothyronine concentrations	Holstein cattle
					Lindholm-Perry et al. (2016)	Feed Intake	beef steers
15	BTB-00590603	RFI	IL10RA*	Immune response	Salleh et al. (2018)	RFI	Jersey cattle
					Gurgul et al. (2019)	Immune response	Polish cattle
					Zhou et al. (2018)	Immune response	Holstein cattle
			TMPRSS4*	Scavenger receptor activity			
17	Hapmap49910-BTA-20754	RFI	IL21**	Immune response	1. Hou et al. (2012)	RFI	Holstein cattle
					Gurgul et al. (2019)	Immune response	Polish cattle

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

4.4.2.4. Analysis between models

Comparative analysis between models was performed. With the significant SNPs at $P < \alpha = 0.001$ in GWAS and quantile effect and importance > 0.999 in Bayes LASSO and RF, respectively.

Coincident SNPs between GWAS and Bayes LASSO

In calves:

- for DMI, three common SNPs were found, these SNPs (and their BTA) were *Hapmap43629-BTA-60810* (1), *Hapmap51428-BTA-26864* (1) and *BTA-86837-no-rs* (20).
- For RFI, two common SNPs were found: *Hapmap41613-BTA-67108* (19) and *BTA-86837-no-rs* (20), the last one was also coincident with DMI.
- There were no common SNPs between FCE and ADG.

In cows:

- For DMI, five common SNPs (BTA) were found, these were *ARS-BFGL-NGS-29072* (9), *BTB-01286081* (9), *BTB-00393938* (9), *Hapmap48321-BTA-40830* (17) and *ARS-BFGL-NGS-18633* (28); of which, the second and third one were at a distance of approximately 62 kb.
- For FCE , ten SNPs (BTA) were found, these were *BTB-01195369* (3), *BTB-00493207* (12), *ARS-BFGL-BAC-13721* (12), *Hapmap54034-rs29026486* (13), *BTA-28181-no-rs* (13), *ARS-BFGL-NGS-61425* (15), *ARS-BFGL-NGS-110736* (15), *ARS-BFGL-BAC-33541* (15), *Hapmap48340-BTA-43615* (18) and *ARS- BFGL-NGS-111019* (28); of which, the seventh and eighth one were located at a distance of approximately 70 kb.
- Eight SNPs (BTA) were found for RFI, these were *BTB-00035766* (1), *BTB-01141030* (2), *UA-IFASA-4367* (8), *ARS-BFGL-NGS-85644* (8), *BTB-00631737* (8), *BTB-00631715* (8), *Hapmap34677-BES4_Contig489_1116* (8) and *Hapmap26379-BTA-130999* (16); of which, the fifth and sixth one were located at a distance of approximately 24 kb.
- Seven SNPs (BTA) were found for ROFC, these were *Hapmap45649-BTA-29691* (2), *ARS-BFGL-NGS-118243* (3), *Hapmap35523-SCAFFOLD5083_24631* (13), ARS-BFGL-NGS-61425 (15), *ARS-BFGL-NGS-82204* (19), *ARS-BFGL-NGS-118018* (19)

and *ARS-BFGL-NGS-109844* (19); of which, the distance between the last three was less than 200 kb; besides, the marker *ARS-BFGL-NGS-61425* was also coincident for FCE.

The common markers between these two models in cows was larger than in calves, probably because the association in calves had a smaller sample size (n = 30).

Common SNPs between GWAS and random forest

In calves:

- for DMI, one common SNPs (BTA) was found, this was *BTB-00031796* (1).
- For FCE no common SNPs were found.
- For RFI, one common SNP was found, this was *ARS-BFGL-NGS-116361* (1).
- For ADG, two common SNPs were found, these were *Hapmap49963-BTA-33040* (13) and *ARS-BFGL-NGS-21830* (13), and they were located at approximately 200 kb apart.

In cows:

- For DMI, one common SNP (BTA) was found, this was *ARS-BFGL-NGS-89583* (11).
- There were no common SNPs for FCE.
- For RFI, nine common SNPs (BTA) were found, these were *ARS-BFGL-NGS-17993* (8), *BTB-00631737* (8), *BTB-00631715* (8), *Hapmap34677-BES4_Contig489_1116* (8), *ARS-BFGL-NGS -117511* (9), *ARS-BFGL-NGS-22941* (10), *ARS-BFGL-NGS-63663* (13), *Hapmap49910-BTA-20754* (17) and *Hapmap39026-BTA-42843* (18), of which the second and third were located at approximately 24 kb apart.
- For ROFC, two common SNPs (BTA) were found, these were *ARS-BFGL-NGS-118018* (19) and *ARS-BFGL-NGS-109844* (19), the two were located at approximately 206 kb apart.

In this section, it can also be observed that there were more common SNPs between models for cows than for calves.

Common SNP between Bayes LASSO and random forest

In calves:

- Only one SNP in common was found for DMI, this SNPs (BTA) was *BTB-00029666* (1).
- For the other FE traits, there was no coincidence of SNPs.

In cows:

- None common SNP (BTA) was found for DMI.
- For FCE, three common SNPs were found, these were *Hapmap53424-rs29019267* (3), *ARS-BFGL-NGS-82204* (19) and *ARS-BFGL-NGS-118018* (19); the last two were at approximately 58 kb.
- For RFI, four common SNPs (BTA) were found, these were *BTB-00631737* (8), *BTB-00631715* (8), *Hapmap34677-BES4_Contig489_1116* (8) and *ARS-BFGL-NGS-3005* (10); of which, the first and second one were located at approximately 24 kb between them.
- For ROFC, two common SNPs were found, these were *ARS-BFGL-NGS-118018* (19) and *ARS-BFGL-NGS-109844* (19), these were located at approximately 106 kb between them; besides, *ARS-BFGL-NGS-118018* was also significant for FCE.

As in the two previous comparisons, cows presented more coincidence of SNPs between models than calves. These results were summarized in Figure 15.

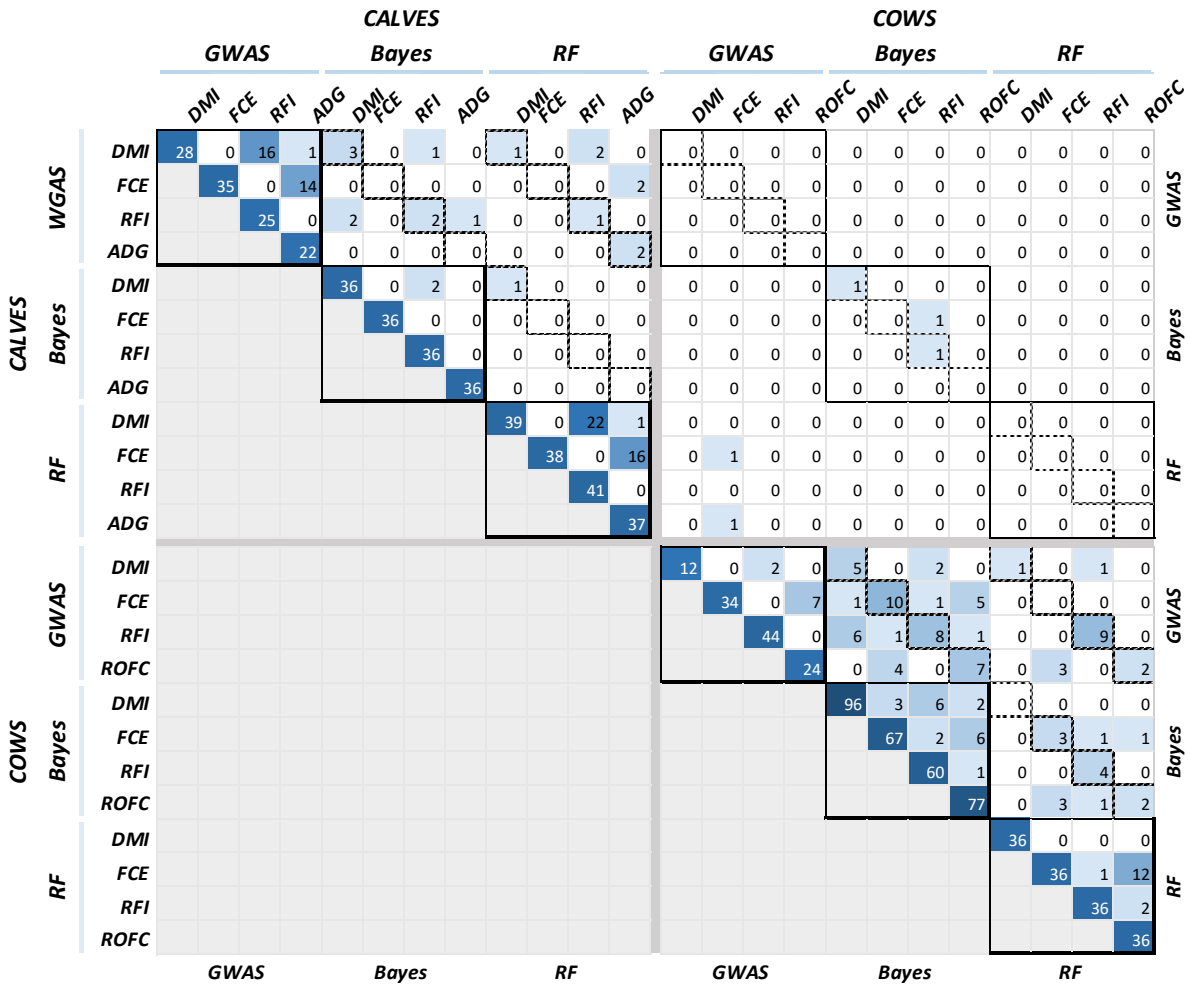


Figure 15. SNPs in common between models and between groups. SNPs with alpha <0.001 and quantile > 0.999 are shown in blue. In the upper-right and lower-left corner, the SNPs in common between cows and calves are shown.

In a general, the common SNPs between models was scarce in calves, while in cows the GWAS and Bayes LASSO models share more SNPs in common. Furthermore, the SNPs shared between cows and calves was practically null and only in the Bayes LASSO, two shared SNPs (BTA) were found, these were:

- *Hapmap48117-BTA-90454,* this was located on chromosome 12, the closest gene was ENSBTAG00000053445 (approximately 200kb away) and its associated biological process was the negative regulation of phosphatase activity.
- *BTA-43831-no-rs,* this was found on chromosome 18, it was found within the LOC785907 gene and its associated molecular function is symporter activity.

SNPs in common in the three statistical models

SNPs were selected at alpha <0.001, effect at quantile > 0.999 and importance at quantile > 0.999 for GWAS, BL and RF, respectively. Two points were reported as relevant.

- Only three common SNPs for RFI and two for ROFC were recurrent in cows in the three statistical models (detailed in Table 17) These genes have not been found in other association studies related to food production or efficiency.
- When SNPs were compared between cows and calves, none were common between FE traits.

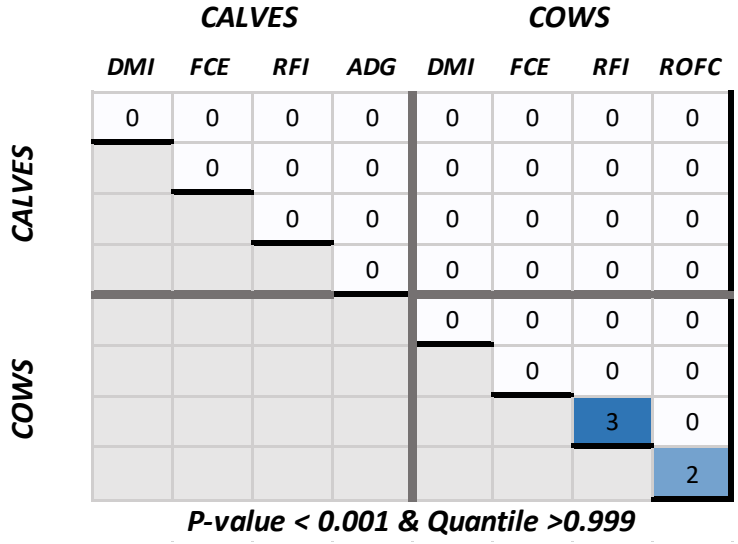


Figure 16. Common SNPs in the three models, with alpha <0.001 and quantile 0.999. The upper right quadrant shows the common SNPs between cows and calves.

Table 17. SNPs in common in the three statistical models.

Ch	Marker	Trait	QTL	Associated function
8	Hapmap34677-BES4_Contig489_1116	RFI	FAM205C**	Protein coding
			ENSBTAG00000050015*	Peptidyl-prolyl cis-trans isomerase activity
	BTB-00631715	RFI	FOXB2*	Anatomical structure
	BTB-00631737	RFI	FOXB2*	Anatomical structure
19	ARS-BFGL-NGS-118018	ROFC	ALKBH5	cell differentiation mRNA processing spermatogenesis
	ARS-BFGL-NGS-109844	ROFC	TOM1L2	intracellular protein transport negative regulation of mitotic nuclear division

Ch = chromosome; QTL displays the gene or two genes closest to the marker; * indicates that the gene was within 200kb before or after from the SNP; ** indicates that the gene was more than 200kb away from the SNP; when it does not present * it means that the SNP was within the gene.

4.5. Genome wide prediction

As mentioned above (item 3.6); predictions in cows were implemented using genomic information from calves and their phenotypes. To evaluate the accuracy of the results, two evaluations were developed, 1) the correlation between the cows GEBV and their phenotypes (corrected by systematic effects) and 2) the mean squared error (MSE) between the cows GEBV and their corrected phenotype.

It should be remembered that the phenotypes used in the GBLUP were rescaled between 0.02 and 1 (the same ones used in these evaluations). The results are detailed below in Table 18 and Figure 17 and Figure 18.

4.5.1. Accuracy of prediction

Table 18. Correlation and MSE from genomic prediction of feed efficiency in cows

	r	*MSE*
DMI	-0.00484	0.3262046
FCE	-0.0705	0.2428647
RFI	0.0421	0.3567229

r = correlation between the GEBV and phenotype (corrected by systematic effects) in cows; MSE = mean squared error between the GEBV and scaled phenotype (corrected by systematic effects) in cows.

Correlations between GEBV and phenotype:

Genomic predictions showed unfavourable accuracy for cows; the correlation between cows GEBV and their corrected phenotype were -0.005, -0.071, and -0.042 for DMI, FCE, and RFI, respectively. Figure 17 show that, cow GEBV (estimated with the genotype and phenotype from calves) were no related with their phenotype. These results suggest that FE traits in calves were different from their homologous in adult cows.

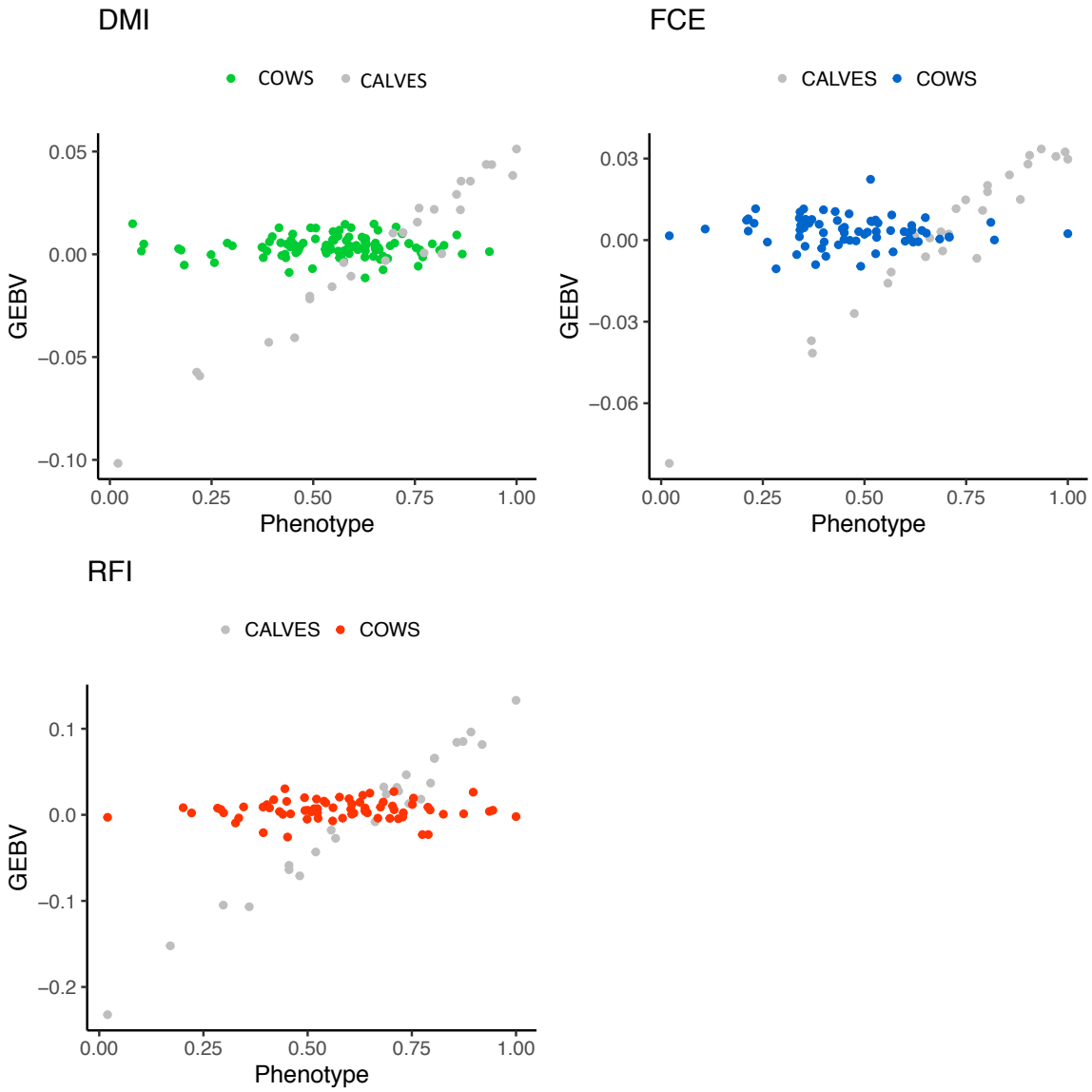


Figure 17. Scatter plots for GEBV (with calves genomic data as proxies) with scaled and corrected phenotype. In grey colour the relation between the calves GEBV with their phenotype is observed, these values make a visual reference for better visualization of the results in cows.

These studies were not conducted in early calves, and do not coincide with other prediction with heifers; Davis et al., (2014) showed that it was possible to discriminate by RFI in lactating cows from the estimated GEBV in growing heifers. Macdonald et al. (2014b) conclude that calves diverging for RFI during growth (6-9 months of age) were also divergent for RFI during lactation and although the difference in lactation was small, this was statistically significant. The calves in this study were in early life-stages, so that ruminal development was limited, and this could explain the genetic difference of FE traits between cows and calves.

Correlations between GEBV and phenotype:

Without knowing the true genetic value, the MSE was estimated using the predicted genetic value (GEBV) and their corrected phenotype. The results showed unfavourable accuracies. Figure 18 shows the phenotype, the GEBV and the MSE. When comparing these three boxplots, the GEBV had values close to zero and their variance was reduced. Furthermore, the MSE was very high which suggests that the genes controlling the expression of FE in calves cannot explain the genetic expression of FE in cows.

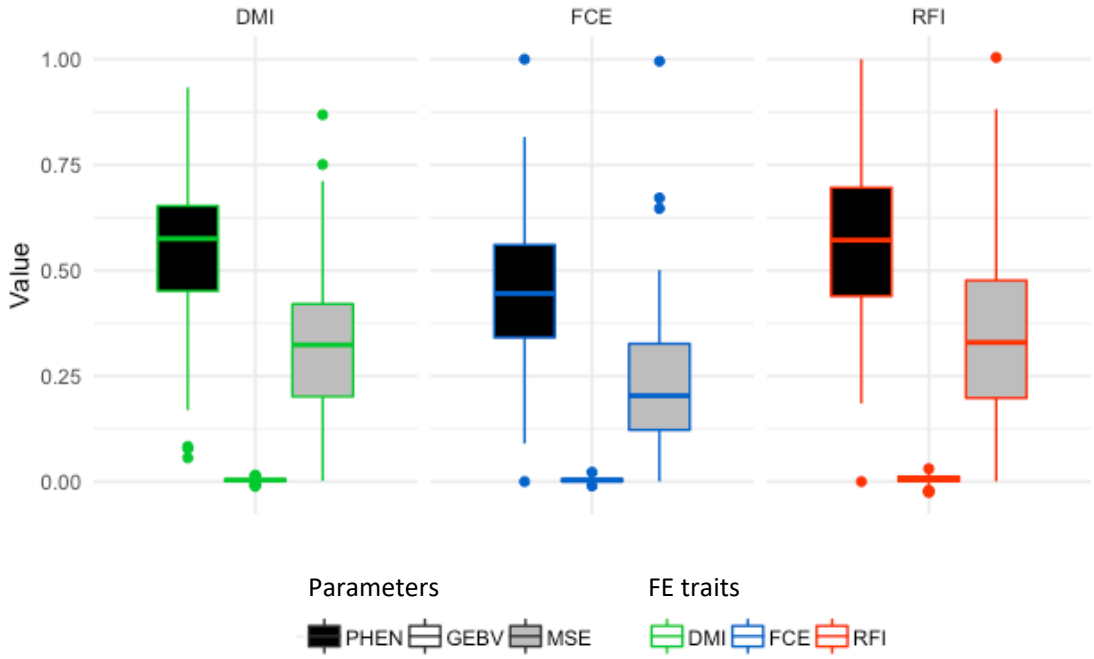


Figure 18. Boxplot of Phenotype (scaled and corrected by systematic effects), GEBV (predicted with proxies from calves), and MSE (between phenotype and GEBV) of FE in cows.

5. CONCLUSIONS

5. CONCLUSIONS

Phenotypic correlations of FE traits

- In general, feed efficiency related traits showed strong phenotypic correlations between them, although not close to 1, suggesting that each trait measures different FE aspects. Stronger correlation was found in calves than in cows, suggesting that FE in calves is biologically less complex.

Phenotypic prediction of FE traits from proxies in cows

- Predicting FE in cows is cumbersome, but a moderate predictive accuracy is possible applying the right models. The best predictor models were Bagging (r = 0.58; MSE = 3.52) for DMI, a linear model for FCE (r = 0.85; MSE = 0.018), and NNET for ROFC (r = 0.99; MSE = 0.13).

Variance component estimations for FE traits.

- Heritabilities were high and their standard deviation were moderate for all traits, although slightly larger estimates were found in calves, suggesting that FE in calves is less influenced by the environment.
- Genetic correlations between FE traits in calves were higher than in cows, although they were estimated with large uncertainty.

Genomic analysis involved in calf and cow FE.

- The GWAS showed few common SNPs associated with several traits in calves or cows. The genes close to the most significant SNPs involved many biological functions, although, the genes associated with FE showed relationship with immunity response, ion transport, cell development and enzyme functions.
- Only one SNP for DMI and one SNP for RFI were shared between cows and calves. These were *Hapmap48117-BTA-90454* and *BTA-43831-no-rs,* their nearest genes were ENSBTAG00000053445 and LOC785907, and their associated functions were the negative regulation of phosphatase activity and symporter activity, respectively
- Results in this thesis suggest that FE traits in cows and calves are regulated by different genes.

Genomic prediction of FE traits from early life proxies.

- Accuracies of genomic predictions for FE traits were low, the correlations between GEBV and corrected phenotype were -0.005, -0.071 and -0.042 for DMI, FCE and RFI, respectively. Therefore, it was not possible to predict FE traits from early life proxies. This emphasizes that FE is controlled by different genes in cows and calves.

General conclusion.

- Based on the results obtained in this study, it seems that the statistical genetic architecture that controls FE in calves and cows is different, making it difficult to use information from calves to predict FE in adult cows. Although the sample size of this study is small and the strength of the evidence is low, it is suggested to verify these results in a larger data set.

6. CITED LITERATURE

6. CITED LITERATURE

Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet. Sel. Evol. 51:28. doi:10.1186/s12711-019-0469-3.

An, B., L. Xu, J. Xia, X. Wang, J. Miao, T. Chang, M. Song, J. Ni, L. Xu, L. Zhang, J. Li, and H. Gao. 2020. Multiple association analysis of loci and candidate genes that regulate body size at three growth stages in Simmental beef cattle. BMC Genet. 21. doi:10.1186/s12863-020-0837-6.

Archer, J.A., A. Reverter, R.M. Herd, D.J. Johnston, and P.F. Arthur. 2002. Genetic Variation in Feed Intake and Efficiency of Mature Feef Cows and Relationships with Postweaning Measurements. Accessed. https://www.researchgate.net/profile/Jason_Archer/publication/266573238_Genetic_variation_in_feed_intake_and_efficiency_of_mature_beef_cows_and_relationships_with_postweaning_measurements/links/5542f2dd0cf24107d3948ca0.pdf.

Van Arendonk, J.A.M., G.J. Nieuwhof, H. Vos, and S. Korver. 1991. Genetic aspects of feed intake and efficiency in lactating dairy heifers. Livest. Prod. Sci. 29:263–275. doi:10.1016/0301-6226(91)90103-W.

Arthur, P.F., J.A. Archer, D.J. Johnston, R.M. Herd, E.C. Richardson, and P.F. Parnell. 2001a. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805–2811. doi:10.2527/2001.79112805x.

Arthur, P.F., G. Renand, and D. Krauss. 2001b. Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. Livest. Prod. Sci. 68:131–139. doi:10.1016/S0301-6226(00)00243-8.

Bach, A. 2014. Precision Feeding to Increase Efficiency for Milk Production. Accessed. https://wcds.ualberta.ca/wcds/wp-content/uploads/sites/57/wcds_archive/Archive/2014/Manuscripts/p 177 - 192 Bach.pdf.

Bach, À. 2005. El Consumo de Materia Seca En El Vacuno Lechero. Accessed November 25,

2019. http://62.174.80.130/articulos/n174/A17403.pdf.

Ball, R. 2013. Package 'LdDesign.' Accessed March 25, 2020. https://github.com/cran/ldDesign/blob/master/DESCRIPTION.

Beever, D.E., and P.T. Doyle. 2007. Feed conversion efficiency as a key determinant of dairy herd performance: A review. Aust. J. Exp. Agric. 47:645–657. doi:10.1071/EA06048.

Berry, D.P., and J.J. Crowley. 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. J. Anim. Sci. 91:1594–1613. doi:10.2527/jas.2012-5862.

Beygelzimer, A., S. Kakadet, J. Langford, S. Arya, D. Mount, and L. Shengqiao. 2019. Package "FNN." Accessed March 24, 2020. https://cran.r-project.org/web/packages/FNN/FNN.pdf.

Biffani, S., N. Morandi, V. Locatelli, D. Pravettoni, A. Boccardo, A. Stella, E.L. Nicolazzi, and F. Biscarini. 2014. Adding evidence for a role of the SLITRK gene family in the pathogenesis of left displacement of the abomasum in Holstein-Friesian dairy cows. Livest. Sci. 167:104–109. doi:10.1016/j.livsci.2014.05.002.

Blasco, A. 2017. Bayesian Data Analysis for Animal Scientists. Springer, Cham, Switzerland.

Blasco, A., and M.A. Toro. 2014. A short critical history of the application of genomics to animal breeding. Livest. Sci. 166:4–9. doi:10.1016/j.livsci.2014.03.015.

Bolormaa, S., B.J. Hayes, K. Savin, R. Hawken, W. Barendse, P.F. Arthur, R.M. Herd, and M.E. Goddard. 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89:1684–1697. doi:10.2527/jas.2010-3079.

Boyle, E.A., Y.I. Li, and J.K. Pritchard. 2017. Leading edge perspective an expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186. doi:10.1016/j.cell.2017.05.038.

Bozic, M., J. Newton, C.S. Thraen, and B.W. Gould. 2012. Mean-reversion in income over feed cost margins: Evidence and implications for managing margin risk by US dairy producers. J. Dairy Sci. 95:7417–7428. doi:10.3168/jds.2012-5818.

Breiman, L. 1996. Bagging predictors. Mach. Learn. 24:123–140. doi:10.1007/bf00058655.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32. doi:10.1023/A:1010933404324.

Browning, B.L., and S.R. Browning. 2008. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210–223. doi:10.1016/j.ajhg.2009.01.005.

Buitenhuis, B., L.L.G. Janss, N.A. Poulsen, L.B. Larsen, M.K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi:10.1186/1471-2164-15-1112.

Cai, Z., M. Dusza, B. Guldbrandtsen, M.S. Lund, and G. Sahana. 2020. Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. Genet. Sel. Evol. 52. doi:10.1186/s12711-020-00538-6.

Carvalho, F.E., R. Espigolan, M.P. Berton, J.B.S. Neto, R.P. Silva, L. Grigoletto, R.M.O. Silva, J.B.S. Ferraz, J.P. Eler, I. Aguilar, R.B. Lôbo, and F. Baldi. 2020. Genome-wide association study and predictive ability for growth traits in Nellore cattle. Livest. Sci. 231:103861. doi:10.1016/j.livsci.2019.103861.

Charfeddine, N., and M.A. Pérez-Cabal. 2019. Estudio Económico Para La Actualización Del ICO y Definición de Nuevos Índices de Mérito Económico Total. Accessed. http://www.revistafrisona.com/Noticia/estudio-economico-para-la-actualizacion-del-ico-y-definicion-de-nuevos-indices-de-merito-economico-total.

Clark, S.A., and J. Van Der Werf. 2013. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. Methods Mol. Biol. 1019:321–330. doi:10.1007/978-1-62703-447-0_13.

Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor, B.A. Crooker, C.P. Van Tassell, J. Yang, S. Wang, L.K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 12:1–17. doi:10.1186/1471-2164-12-408.

Connor, E.E. 2015. Invited review: Improving feed efficiency in dairy production: Challenges and possibilities. Animal 9:395–408. doi:10.1017/S1751731114002997.

Cottle, D., and L. Kahn. 2014. Beef Cattle Production and Trade. CSIRO, Victoria, Australia.

Cue, R.I., H.G. Monardes, and J.F. Hayes. 1987. Correlations between production traits in first lactation holstein cows. J. Dairy Sci. 70:2132–2137. doi:10.3168/jds.S0022-0302(87)80264-3.

Davis, S.R., K.A. Macdonald, G.C. Waghorn, and R.J. Spelman. 2014. Residual feed intake of lactating Holstein-Friesian cows predicted from high-density genotypes and phenotyping of growing heifers. J. Dairy Sci. 97:1436–1445. doi:10.3168/jds.2013-7205.

Dechow, C.D., G.W. Rogers, and J.S. Clay. 2002. Heritability and correlations among body condition score loss, body condition score, production and reproductive performance. J. Dairy Sci. 85:3062–3070. doi:10.3168/jds.S0022-0302(02)74393-2.

Delgado, B., A. Bach, I. Guasch, C. González, G. Elcoso, J.E. Pryce, and O. Gonzalez-Recio. 2019. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. Sci. Rep. 9. doi:10.1038/s41598-018-36673-w.

Deng, T., C. Pang, X. Ma, X. Lu, A. Duan, P. Zhu, and X. Liang. 2016. Four novel polymorphisms of buffalo INSIG2 gene are associated with milk production traits in Chinese buffaloes. Mol. Cell. Probes 30:294–299. doi:10.1016/j.mcp.2016.09.003.

Duncombe, J.L. 2016. An Evaluation of Gene Interactions Affecting Carcass Yield and Marbling in Beef Cattle. Accessed May 26, 2020. https://harvest.usask.ca/bitstream/handle/10388/7369/DUNCOMBE-THESIS-2016.pdf?sequence=1&isAllowed=y.

El-Halawany, N., H. Abdel-Shafy, A.E.M.A. Shawky, M.A. Abdel-Latif, A.F.M. Al-Tohamy, and O.M. Abd El-Moneim. 2017. Genome-wide association study for milk production in Egyptian buffalo. Livest. Sci. 198:10–16. doi:10.1016/j.livsci.2017.01.019.

Elkan, C. 2011. Evaluating Classifiers. Accessed April 20, 2020. https://web.archive.org/web/20111218192652/http://cseweb.ucsd.edu/~elkan/250B/classifiereval.pdf.

Fisher, R.A. 1919. XV.—The Correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinburgh 52:399–433. doi:10.1017/S0080456800012163.

Fix, E., and J.L. Hodges. 1989. Discriminatory analysis. nonparametric discrimination: consistency Properties. Int. Stat. Rev. / Rev. Int. Stat. 57:238. doi:10.2307/1403797.

Fritsch, S., F. Guenther, M.N. Wright, M. Suling, and S.M. Mueller. 2019. Package 'Neuralnet.' Accessed March 24, 2020. https://github.com/bips-hb/neuralnet BugReports.

Gan, Q.F., Y.R. Li, Q.H. Liu, M. Lund, G.S. Su, and X.W. Liang. 2019. Genome-wide association studies for the concentrations of insulin, triiodothyronine, and thyroxine in Chinese Holstein cattle. Trop. Anim. Health Prod. 1–6. doi:10.1007/s11250-019-02170-z.

Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124:323–330. doi:10.1111/jbg.2007.124.issue-6.

Gonçalves, T.M., L.C. De Almeida Regitano, J.E. Koltes, A.S.M. Cesar, S.C. Da Silva Andrade, G.B. Mourão, G. Gasparin, G.C.M. Moreira, E. Fritz-Waters, J.M. Reecy, and L.L. Coutinho. 2018. Gene co-expression analysis indicates potential pathways and regulators of beef tenderness in Nellore cattle. Front. Genet. 9:441. doi:10.3389/fgene.2018.00441.

Gondro, C., J. van der Werf, and B. Hayes. 2013. Genome-Wide Association Studies and Genomic Prediction. Springer, Totowa, NJ.

Gonzalez-Recio, O. 2010. RanFoG, random forest for classified and regression problems.

González-Recio, O., and S. Forni. 2011. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. Genet. Sel. Evol. 43:7. doi:10.1186/1297-9686-43-7.

Gonzalez-Recio, O., J.E. Pryce, M. Haile-Mariam, and B.J. Hayes. 2014. Incorporating heifer feed efficiency in the Australian selection index using genomic selection. J. Dairy Sci. 97:3883–3893. doi:10.3168/jds.2013-7515.

González-Recio, O., G.J.M. Rosa, and D. Gianola. 2014. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest. Sci. 166:217–231. doi:10.1016/j.livsci.2014.05.036.

Green, T.C., J.G. Jago, K.A. Macdonald, and G.C. Waghorn. 2013. Relationships between residual feed intake, average daily gain, and feeding behavior in growing dairy heifers. J. Dairy Sci. 96:3098–3107. doi:10.3168/jds.2012-6087.

Grobet, L., L.J.R. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S.

Dunner, F. Ḿnissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17:71–74. doi:10.1038/ng0997-71.

Guo, P., Z. Zhao, S. Yan, J. Li, H. Xiao, D. Yang, Y. Zhao, P. Jiang, and R. Yang. 2016. PSAP gene variants and haplotypes reveal significant effects on carcass and meat quality traits in Chinese Simmental-cross cattle. Arch. Anim. Breed 59:461–468. doi:10.5194/aab-59-461-2016.

Gurgul, A., I. Jasielczuk, E. Semik-Gurgul, T. Szmatoła, A. Majewska, E. Sosin-Bzducha, and M. Bugno-Poniewierska. 2019. Diversifying selection signatures among divergently selected subpopulations of Polish Red cattle. J. Appl. Genet. 60:87–95. doi:10.1007/s13353-019-00484-0.

De Haas, Y., J.E. Pryce, M.P.L. Calus, E. Wall, D.P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K.A. Macdonald, B. Hulsegge, and R.F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. J. Dairy Sci. 98:6522–6534. doi:10.3168/jds.2014-9257.

Hall, T. 2011. Define And Improve Your Herd's Feed Conversion Ratio. Accessed November 25, 2019. https://web.archive.org/web/20160616035304/http://lallemandanimalnutrition.com/wp-content/uploads/2011/10/Hal_Dairybusiness_09111.pdf.

Hardie, L.C., M.J. VandeHaar, R.J. Tempelman, K.A. Weigel, L.E. Armentano, G.R. Wiggans, R.F. Veerkamp, Y. de Haas, M.P. Coffey, E.E. Connor, M.D. Hanigan, C. Staples, Z. Wang, J.C.M. Dekkers, and D.M. Spurlock. 2017. The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. J. Dairy Sci. 100:9061–9075. doi:10.3168/jds.2017-12604.

Hegarty, R.S., J.P. Goopy, R.M. Herd, and B. McCorkell. 2007. Cattle selected for lower residual feed intake have reduced daily methane production. J. Anim. Sci. 85:1479–1486. doi:10.2527/jas.2006-236.

Hempstalk, K., S. McParland, and D.P. Berry. 2015. Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. J. Dairy

Sci. 98:5262–5273. doi:10.3168/jds.2014-8984.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423. doi:10.2307/2529430.

Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69. doi:10.2307/2529339.

Hoque, M.A., P.F. Arthur, K. Hiramoto, A.R. Gilmour, and T. Oikawa. 2007. Variance components due to direct genetic, maternal genetic and permanent environmental effect for growth and feed-efficiency traits in young male Japanese Black cattle. J. Anim. Breed. Genet. 124:102–107. doi:10.1111/j.1439-0388.2007.00648.x.

Hou, Y., D.M. Bickhart, H. Chung, J.L. Hutchison, H.D. Norman, E.E. Connor, and G.E. Liu. 2012. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct. Integr. Genomics 12:717–723. doi:10.1007/s10142-012-0295-y.

Hutjens, M.F. 2012. Feed Efficiency and Its Impact on Feed Intake. Accessed November 26, 2019. https://web.archive.org/web/20160616043236/http://articles.extension.org/pages/11317/feed-efficiency-and-its-impact-on-feed-intake.

Irano, N., G.M.F. De Camargo, R.B. Costa, A.P.N. Terakado, A.F.B. Magalhães, R.M.D.O. Silva, M.M. Dias, A.B. Bignardi, F. Baldi, R. Carvalheiro, H.N. De Oliveira, and L.G. De Albuquerque. 2016. Genome-wide association study for indicator traits of sexual precocity in Nellore cattle. PLoS One 11. doi:10.1371/journal.pone.0159502.

Jiménez-Montero, J.A., O. González-Recio, and R. Alenda. 2013. Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. J. Dairy Sci. 96:625–634. doi:10.3168/jds.2012-5631.

Jong, G. De, R. Van Der Linde, Y. De Haas, G.C.B. Schopen, and R.F. Veerkamp. 2016. Genetic Evaluation for Feed Intake in the Netherlands and Flanders, Impact on Efficiency and Responses. Accessed. https://journal.interbull.org/index.php/ib/article/view/1390.

Jorge Matich, D. 2001. Redes Neuronales: Conceptos Básicos y Aplicaciones. Accessed March 24, 2020. https://docplayer.es/8032554-Catedra-informatica-aplicada-a-la-ingenieria-de-

procesos-orientacion-i-tema-redes-neuronales-conceptos-basicos-y-aplicaciones.html.

Júnior, G.A.F., R.B. Costa, G.M.F. de Camargo, R. Carvalheiro, G.J.M. Rosa, F. Baldi, D.A. Garcia, D.G.M. Gordo, R. Espigolan, L. Takada, A.F.B. Magalhães, T. Bresolin, F.L.B. Feitosa, L.A.L. Chardulo, H.N. de Oliveira, and L.G. de Albuquerque. 2016. Genome scan for postmortem carcass traits in Nellore cattle. J. Anim. Sci. 94:4087–4095. doi:10.2527/jas.2016-0632.

Kassambara, A., and F. Mundt. 2019. Package 'Factoextra.' Accessed March 28, 2020. http://www.sthda.com/english/rpkgs/factoextra.

Khansefid, M. 2016. Genomic Selection for Residual Feed Intake. Accessed. http://hdl.handle.net/11343/116294.

Khansefid, M., C.A. Millen, Y. Chen, J.E. Pryce, A.J. Chamberlain, C.J. Vander Jagt, C. Gondro, and M.E. Goddard. 2017. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake. J. Anim. Sci. 95:4764–4775. doi:10.2527/jas2016.1320.

Knutsen, T.M., H.G. Olsen, V. Tafintseva, M. Svendsen, A. Kohler, M.P. Kent, and S. Lien. 2018. Unravelling genetic variation underlying de novo-synthesis of bovine milk fatty acids. Sci. Rep. 8:1–13. doi:10.1038/s41598-018-20476-0.

Koch, R.M., L.A. Swiger, D. Chambers, and K.E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494. doi:10.2527/jas1963.222486x.

Korver, S., E.A.M. van Eekelen, H. Vos, G.J. Nieuwhof, and J.A.M. van Arendonk. 1991. Genetic parameters for feed intake and feed efficiency in growing dairy heifers. Livest. Prod. Sci. 29:49–59. doi:10.1016/0301-6226(91)90119-B.

de las Heras-Saldana, S., S.A. Clark, N. Duijvesteijn, C. Gondro, J.H.J. Van Der Werf, and Y. Chen. 2019. Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. BMC Genomics 20:1–16. doi:10.1186/s12864-019-6270-4.

Lee, S., H. Min-Wook, C. So-Young, and J. Kim. 2018. WPSII-5 Genome-wide association analysis to identify QTL for carcass traits in Korean native cattle.. J. Anim. Sci. 96:516. doi:10.1093/JAS/SKY404.1130.

Legarra, A., L. Varona, and E. López de Maturana. 2011. Threshold Model. Accessed March 25, 2020. http://snp.toulouse.inra.fr/~alegarra/manualtm.pdf.

Li, B., W.F. Fikse, J. Lassen, M.H. Lidauer, P. Løvendahl, P. Mäntysaari, and B. Berglund. 2016. Genetic parameters for dry matter intake in primiparous Holstein, Nordic Red, and Jersey cows in the first half of lactation. J. Dairy Sci. 99:7232–7239. doi:10.3168/jds.2015-10669.

Li, C., D. Sun, S. Zhang, S. Wang, X. Wu, Q. Zhang, L. Liu, Y. Li, and L. Qiao. 2014. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. PLoS One 9. doi:10.1371/journal.pone.0096186.

Li, Y., Y. Gao, Y.S. Kim, A. Iqbal, and J.J. Kim. 2017. A whole genome association study to detect additive and dominant single nucleotide polymorphisms for growth and carcass traits in Korean native cattle, Hanwoo. Asian-Australasian J. Anim. Sci. 30:8–19. doi:10.5713/ajas.16.0170.

De Lima, A.O., P.C. Tizioto, P.S.N. De Oliveira, L. Geistlinger, M.I.P. Rocha, W.J. Da, S. Diniz, J. Afonso, C.E. Buss, J. Petrini, G.B. Mourão, L.L. Coutinho, L.C. De, and A. Regitano. 2017. Potential Regulatory Elements on PCDH7 Gene Affecting Residual Feed Intake in Nelore Cattle. Accessed May 26, 2020. https://www.alice.cnptia.embrapa.br/bitstream/doc/1075636/1/Luciana03.pdf.

Lima, A.O.D., P.S.N. Oliveira, P.C. Tizioto, A.L. Somavilla, W.J.S. Diniz, J.V.D. Silva, S.C.S. Andrade, C. Boschiero, A.S.M. Cesar, M.M. Souza, M.I.P. Rocha, J. Afonso, C.E. Buss, M.A. Mudadu, G.B. Mourao, L.L. Coutinho, and L.C.A. Regitano. 2016. PRUNE2 gene has a potential effect on residual feed intake in Nellore cattle. J. Anim. Sci. 94:152–153. doi:10.2527/jam2016-0318.

Lin, Z., I. Macleod, and J.E. Pryce. 2013. Short communication: estimation of genetic parameters for residual feed intake and feeding behavior traits in dairy heifers.. J. Dairy Sci. 96:2654–2656. doi:10.3168/jds.2012-6134.

Lindholm-Perry, A.K., R.J. Kern, B.N. Keel, W.M. Snelling, L.A. Kuehn, and H.C. Freetly. 2016. Profile of the spleen transcriptome in beef steers with variation in gain and feed Intake. Front. Genet. 7:127. doi:10.3389/fgene.2016.00127.

De Los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and

J.M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182:375–385. doi:10.1534/genetics.109.101501.

Lourenco, D. 2017. Introduction to Genomic Selection. Accessed January 15, 2020. https://www.uscdcb.com/Genotype/cur_density.html.

Lourenco, D., K. Bertrand, H. Bradford, S. Miller, and I. Misztal. 2017. The Promise of Genomics for Beef Improvement. Accessed January 15, 2020. https://www.researchgate.net/profile/Daniela_Lourenco2/publication/321110779_The_Promise_of_Genomics_for_Beef_Improvement/links/5a0e06cbaca27244d285870a/The-Promise-of-Genomics-for-Beef-Improvement.pdf.

Macciotta, N.P.P., G. Gaspa, R. Steri, E.L. Nicolazzi, C. Dimauro, C. Pieramati, and A. Cappio-Borlino. 2010. Using eigenvalues as variance priors in the prediction of genomic breeding values by principal component analysis. J. Dairy Sci. 93:2765–2774. doi:10.3168/jds.2009-3029.

Macdonald, K.A., J.E. Pryce, R.J. Spelman, S.R. Davis, W.J. Wales, G.C. Waghorn, Y.J. Williams, L.C. Marett, and B.J. Hayes. 2014. Holstein-Friesian calves selected for divergence in residual feed intake during growth exhibited significant but reduced residual feed intake divergence in their first lactation. J. Dairy Sci. 97:1427–1435. doi:10.3168/jds.2013-7227.

Maintainer, C., and P. Boutros. 2018. Package "VennDiagram." Accessed July 1, 2020. https://cran.r-project.org/web/packages/VennDiagram/VennDiagram.pdf.

Mallikarjunappa, S., M. Sargolzaei, L.F. Brito, K.G. Meade, N.A. Karrow, and S.D. Pant. 2018. Short communication: Uncovering quantitative trait loci associated with resistance to Mycobacterium avium ssp. paratuberculosis infection in Holstein cattle using a high-density single nucleotide polymorphism panel. J. Dairy Sci. 101:7280–7286. doi:10.3168/jds.2018-14388.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Accessed January 15, 2020. https://www.genetics.org/content/157/4/1819.full.

Mudadu, M.A., L.R. Porto-Neto, F.B. Mokry, P.C. Tizioto, P.S.N. Oliveira, R.R. Tullio, R.T.

Nassu, S.C.M. Niciura, P. Tholon, M.M. Alencar, R.H. Higa, A.N. Rosa, G.L.D. Feijó, A.L.J. Ferraz, L.O.C. Silva, S.R. Medeiros, D.P. Lanna, M.L. Nascimento, A.S. Chaves, A.R.D.L. Souza, I.U. Packer, R.A.A. Torres, F. Siqueira, G.B. Mourão, L.L. Coutinho, A. Reverter, and L.C.A. Regitano. 2016. Genomic structure and marker-derived gene networks for growth and meat quality traits of Brazilian Nelore beef cattle. BMC Genomics 17:1–16. doi:10.1186/s12864-016-2535-3.

Mullen, M.P., D.P. Berry, D.J. Howard, M.G. Diskin, C.O. Lynch, L. Giblin, D.A. Kenny, D.A. Magee, K.G. Meade, and S.M. Waters. 2011. Single Nucleotide Polymorphisms in the Insulin-Like Growth Factor 1 (IGF-1) Gene are Associated with Performance in Holstein-Friesian Dairy Cattle. Front. Genet. 2:3. doi:10.3389/fgene.2011.00003.

National Research Council. 2001. Nutrient Requirements of Dairy Cattle. The National Academies, Washintong, USA.

Neupane, M., T.W. Geary, J.N. Kiser, G.W. Burns, P.J. Hansen, T.E. Spencer, and H.L. Neibergs. 2017. Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. PLoS One 12. doi:10.1371/journal.pone.0188997.

Ngwerume, F., and I.L. Mao. 1992. Estimation of Residual Energy Intake for Lactating Cows Using an Animal Model. J. Dairy Sci. 75:2283–2287. doi:10.3168/jds.S0022-0302(92)77989-2.

Olsen, H.G., T.M. Knutsen, A. Kohler, M. Svendsen, L. Gidskehaug, H. Grove, T. Nome, M. Sodeland, K.K. Sundsaasen, M.P. Kent, H. Martens, and S. Lien. 2017. Genome-wide association mapping for milk fat composition and fine mapping of a QTL for de novo synthesis of milk fatty acids on bovine chromosome 13. Genet. Sel. Evol. 49:1–13. doi:10.1186/s12711-017-0294-5.

Page, B.T., E. Casas, M.P. Heaton, N.G. Cullen, D.L. Hyndman, C.A. Morris, A.M. Crawford, T.L. Wheeler, M. Koohmaraie, J.W. Keele, and T.P.L. Smith. 2002. Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J. Anim. Sci. 80:3077–3085. doi:10.2527/2002.80123077x.

Park, T., and G. Casella. 2008. The Bayesian lasso. J. Am. Stat. Assoc. 103:681–686. doi:10.1198/016214508000000337.

Pearson, T.A., and T.A. Manolio. 2008. How to interpret a genome-wide association study.

JAMA - J. Am. Med. Assoc. 299:1335–1344. doi:10.1001/jama.299.11.1335.

Perez, B.C., B. Santana, and J.P. Eler. 2018. Genome-Wide Association Study for Male and Female Reproductive Performance Indicator Traits in Nellore Cattle. Accessed May 26, 2020. https://www.researchgate.net/publication/332229603.

Perez, P., and G. De Los Campos. 2018. Package 'BGLR.' Accessed March 29, 2020. http://genomics.cimmyt.org/BGLR-extdoc.pdf.

Pérez, P., and G. De Los Campos. 2014. BGLR: A Statistical Package for Whole Genome Regression and Prediction. Accessed March 29, 2020. http://genomics.cimmyt.org/BGLR-extdoc.pdf.

Peters, A., T. Hothorn, B.D. Ripley, and T.T. Beth. 2019. Package "Ipred." Accessed March 24, 2020. https://cran.r-project.org/web/packages/ipred/ipred.pdf.

Plummer, M., N. Best, K. Cowles, K. Vines, D. Sarka, D. Bates, R. Almond, and A. Magnusson. 2019. Package "Coda." Accessed June 1, 2020. https://pbil.univ-lyon1.fr/CRAN/web/packages/coda/coda.pdf.

Pryce, J.E., J. Arias, P.J. Bowman, S.R. Davis, K.A. Macdonald, G.C. Waghorn, W.J. Wales, Y.J. Williams, R.J. Spelman, and B.J. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95:2108–2119. doi:10.3168/jds.2011-4628.

Pryce, J.E., O. Gonzalez-Recio, G. Nieuwhof, W.J. Wales, M.P. Coffey, B.J. Hayes, and M.E. Goddard. 2015. Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. J. Dairy Sci. 98:7340–7350. doi:10.3168/jds.2015-9621.

Pryce, J.E., O. Gonzalez-Recio, J.B. Thornhill, L.C. Marett, W.J. Wales, M.P. Coffey, Y. de Haas, R.F. Veerkamp, and B.J. Hayes. 2014a. Short communication: Validation of genomic breeding value predictions for feed intake and feed efficiency traits. J. Dairy Sci. 97:537–542. doi:10.3168/jds.2013-7376.

Pryce, J.E., W.J. Wales, Y. De Haas, R.F. Veerkamp, and B.J. Hayes. 2014b. Genomic selection for feed efficiency in dairy cattle. Animal 8:1–10. doi:10.1017/S1751731113001687.

R Core Team and contributors worldwide. 2018. Package 'STAT.' Accessed March 24, 2020. https://cran.r-project.org/web/packages/STAT/STAT.pdf.

R Core Team and contributors worldwide. 2019. The R Base Package. Accessed March 29, 2020. https://www.rdocumentation.org/packages/base/versions/3.6.2.

Ramayo-Caldas, Y., G. Renand, M. Ballester, R. Saintilan, and D. Rocha. 2016. Multi-breed and multi-trait co-association analysis of meat tenderness and other meat quality traits in three French beef cattle breeds. Genet. Sel. Evol. 48:37. doi:10.1186/s12711-016-0216-y.

Rauw, W.M. 2009. Resource Allocation Theory Applied to Farm Animal Production. CABI, Bodmin, UK.

Riedmiller, M., and H. Braun. 1993. Direct adaptive method for faster backpropagation learning: The RPROP algorithm. 1993 IEEE Int. Conf. Neural Networks 586–591. doi:10.1109/icnn.1993.298623.

Ripley, B., B. Venables, D.M. Bates, K. Hornik, A. Gebhardt, and D. Firth. 2019. Package "MASS." Accessed April 1, 2020. http://www.stats.ox.ac.uk/pub/MASS4/.

Robinson, D.L., and V.H. Oddy. 2004. Genetic parameters for feed efficiency, fatness, muscle area and feeding behaviour of feedlot finished beef cattle. Livest. Prod. Sci. 90:255–270. doi:10.1016/j.livprodsci.2004.06.011.

Rolf, M.M., J.F. Taylor, R.D. Schnabel, S.D. McKay, M.C. McClure, S.L. Northcutt, M.S. Kerley, and R.L. Weaber. 2012. Genome-wide association analysis for feed efficiency in Angus cattle. Anim. Genet. 43:367–374. doi:10.1111/j.1365-2052.2011.02273.x.

Roseler D. K., Fox D. G., Pell A. N., and Chase L. E. 1997. Evaluation of alternative equations for prediction of intake for Holstein dairy cows. J. Dairy Sci. 80. doi:10.3168/jds.S0022-0302(97)76009-0.

Sainz, R.D., and P. V Paulino. 2004. Residual Feed Intake. Accessed November 26, 2019. https://escholarship.org/uc/item/9w93f7ks.

Salleh, M.S., G. Mazzoni, J.K. Höglund, D.W. Olijhoek, P. Lund, P. Løvendahl, and H.N. Kadarmideen. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics 18. doi:10.1186/s12864-017-3622-9.

Salleh, S.M., G. Mazzoni, P. Løvendahl, and H.N. Kadarmideen. 2018. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. BMC Bioinformatics 19:1–15. doi:10.1186/s12859-018-2553-z.

Schloerke, B., and J. Crowley. 2020. Package 'GGally.' Accessed March 31, 2020. https://cran.r-project.org/web/packages/GGally/GGally.pdf.

Seymour, D., A. Cã, T. Chud, J. Cant, V. Osborne, C. Baes, F. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. doi:10.3168/jds.2019-17414.

Sherman, E.L., J.D. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S.S. Moore. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J. Anim. Sci. 87:37–45. doi:10.2527/jas.2008-0876.

Shi, L., L. Liu, Z. Ma, X. Lv, C. Li, L. Xu, B. Han, Y. Li, F. Zhao, Y. Yang, and D. Sun. 2019. Identification of genetic associations of ECHS gene with milk fatty acid traits in dairy cattle. Anim. Genet. 50:430–438. doi:10.1111/age.12833.

Shike, D.W. 2013. Beef Cattle Feed Efficiency. Accessed November 26, 2019. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1027&context=driftlessconference.

Soller, M., and J.S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25–33. doi:10.1007/BF00303917.

Sorensen, D., and D. Gianola. 2007. Likelihood of Bayesian, and MCMC Methods in Quantitative Genetics. Crop Science, Madison, USA.

Spurlock, D.M., J.C.M. Dekkers, R. Fernando, D.A. Koltes, and A. Wolc. 2012. Genetic parameters for energy balance, feed efficiency, and related traits in Holstein cattle. J. Dairy Sci. 95:5393–5402. doi:10.3168/jds.2012-5407.

Taiyun Wei, M., A. Taiyun Wei cre, V. Simko aut, M. Levy ctb, Y. Xie ctb, Y. Jin ctb, and J. Zemla ctb. 2017. Package "Corrplot." Accessed March 31, 2020. https://cran.r-project.org/web/packages/corrplot/corrplot.pdf.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58:267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Torres-Vázquez, J.A., J.H.J. van der Werf, and S.A. Clark. 2018. Genetic and phenotypic associations of feed efficiency with growth and carcass traits in Australian Angus cattle. J. Anim. Sci. 96:4521. doi:10.1093/JAS/SKY325.

Tyrrell, H.F., and J.T. Reid. 1965. Prediction of the energy value of cow's milk. J. Dairy Sci. 48:1215–1223. doi:10.3168/jds.S0022-0302(65)88430-2.

Uemoto, Y., M. Takeda, A. Ogino, K. Kurogi, S. Ogawa, M. Satoh, and F. Terada. 2020. Genetic and genomic analyses for predicted methane-related traits in Japanese Black steers. Anim. Sci. J. 91. doi:10.1111/asj.13383.

Vallimont, J.E., C.D. Dechow, J.M. Daubert, M.W. Dekleva, J.W. Blum, C.M. Barlieb, W. Liu, G.A. Varga, A.J. Heinrichs, and C.R. Baumrucker. 2011. Short communication: Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. J. Dairy Sci. 94:2108–2113. doi:10.3168/jds.2010-3888.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Veerkamp, R.F., M.P.L. Calus, G. De Jong, R. Van Der Linde, and Y. De Haas. 2014. Breeding Value for Dry Matter Intake for Dutch Bulls Based on DGV for DMI and BV for Predictors. Accessed May 8, 2020. http://library.wur.nl/WebQuery/wurpubs/fulltext/359020.

Veerkamp, R.F., G.C. Emmans, A.R. Cromie, and G. Simm. 1995. Variance components for residual feed intake in dairy cows. Livest. Prod. Sci. 41:111–120. doi:10.1016/0301-6226(94)00056-D.

Wickham, H., W. Chang, L. Henry, T. Lin Pedersen, K. Takahashi, W. Claus, K. Woo, H. Yutani, and D. Dunnington. 2020. Package 'Ggplot2.' Accessed March 29, 2020. https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf.

Wickham, H., and D. Seidel. 2019. Package 'Scales.' Accessed April 19, 2020. https://cran.r-project.org/web/packages/scales/scales.pdf.

Williams, Y.J., J.E. Pryce, C. Grainger, W.J. Wales, N. Linden, M. Porker, and B.J. Hayes. 2011. Variation in residual feed intake in Holstein-Friesian dairy heifers in southern

Australia. J. Dairy Sci. 94:4715–4725. doi:10.3168/jds.2010-4015.

Wu, X., B. Guldbrandtsen, M.S. Lund, and G. Sahana. 2016. Association analysis for feet and legs disorders with whole-genome sequence variants in 3 dairy cattle breeds. J. Dairy Sci. 99:7221–7231. doi:10.3168/jds.2015-10705.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569. doi:10.1038/ng.608.

Yao, C., D.M. Spurlock, L.E. Armentano, C.D. Page, M.J. VandeHaar, D.M. Bickhart, and K.A. Weigel. 2013. Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. J. Dairy Sci. 96:6716–6729. doi:10.3168/jds.2012-6237.

Zamani, P., S. Reza Miraei-Ashtiani, and H. Mohammadi. 2008. Genetic Parameters of Residual Energy Intake and Its Correlations with Other Traits in Holstein Dairy Cattle. Accessed October 9, 2019. http://journals.tubitak.gov.tr/veterinary/issues/vet-08-32-4/vet-32-4-4-0610-18.pdf.

Zhang, X., Q. Chu, G. Guo, G. Dong, X. Li, Q. Zhang, S. Zhang, Z. Zhang, and Y. Wang. 2017. Genome-wide association studies identified multiple genetic loci for body size at four growth stages in Chinese Holstein cattle. PLoS One 12. doi:10.1371/journal.pone.0175971.

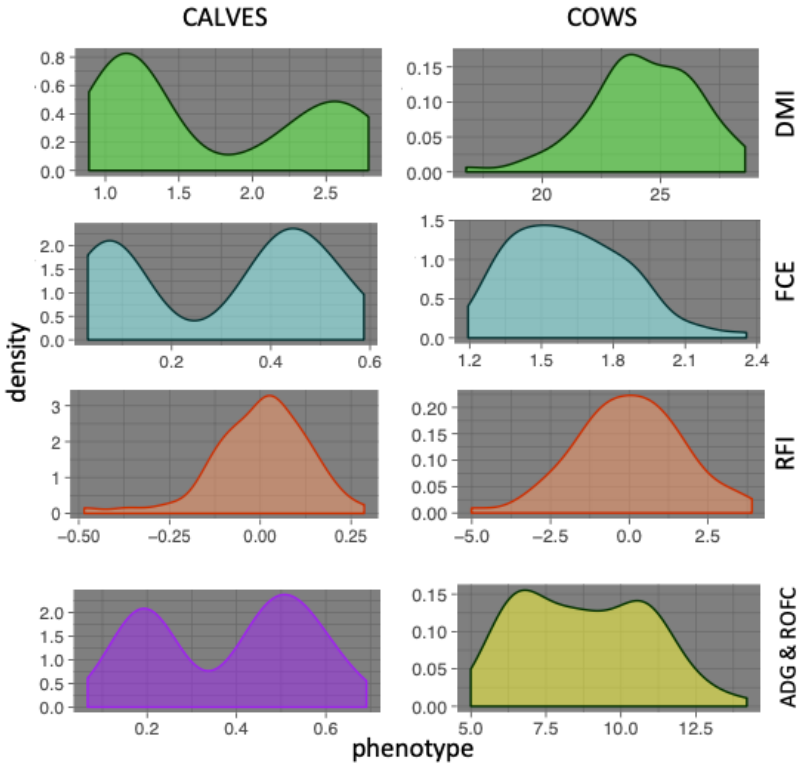
Zhou, J., L. Liu, C.J. Chen, M. Zhang, X. Lu, Z. Zhang, X. Huang, and Y. Shi. 2019. Genome-wide association study of milk and reproductive traits in dual-purpose Xinjiang Brown cattle. BMC Genomics 20:1–11. doi:10.1186/s12864-019-6224-x.

Zhou, Z., F. Ferdous, P. Montagner, D.N. Luchini, M.N. Corrêa, and J.J. Loor. 2018. Methionine and choline supply during the peripartal period alter polymorphonuclear leukocyte immune response and immunometabolic gene expression in Holstein cows. J. Dairy Sci. 101:10374–10382. doi:10.3168/jds.2018-14972.

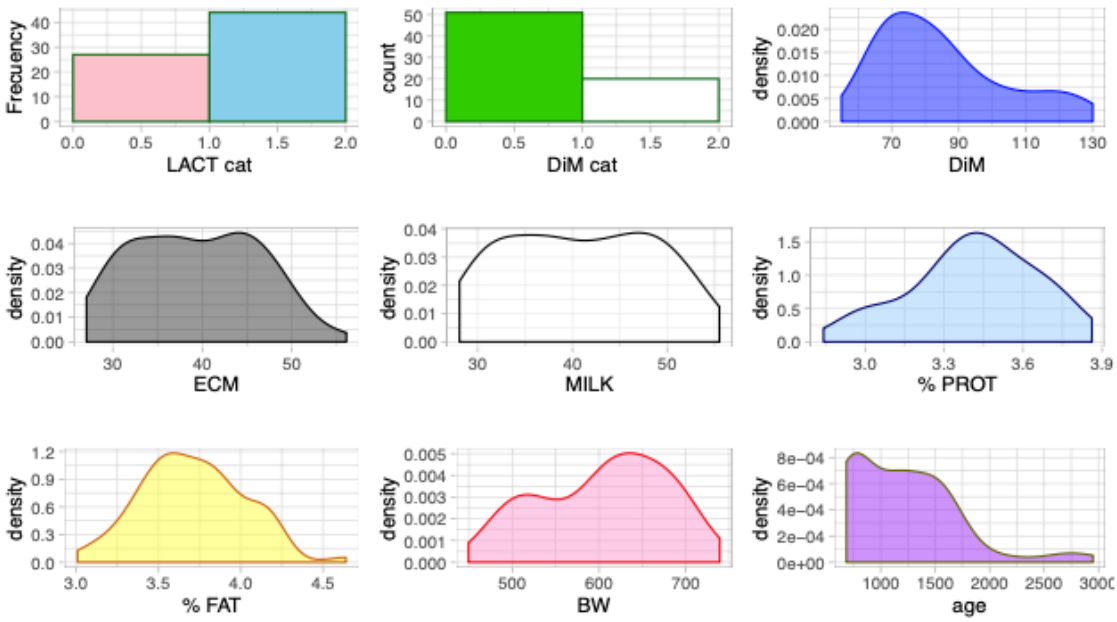
7. ANNEXES

7. ANNEXES

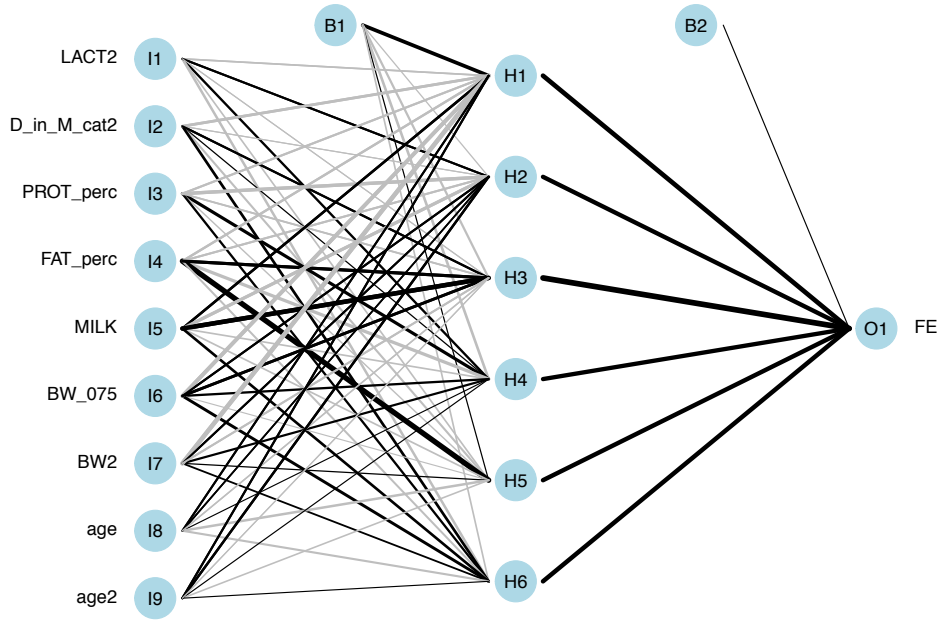
Annex 1. FE-traits distribution.



Annex 2. Productive traits distribution in cows.



Annex 3. Neural networks model for FE in cows.



Annex 4. t test for determinate the difference between predictor models from phenotypic prediction. P adjust method: holm. Phenotypic predictions of FE traits were development using four predictive models. The accuracy of predictions was evaluated through cross validation using the Pearson correlation between real and predicted value. These annexes show the test development to measure the statistical difference between models in the FE traits.

1. p-value for statistical difference between predictive models for DMI

	Bagging	**K-NN**	**LM**
K-NN	0.00149	--	--
LM	0.762	0.07223	--
NNET	1.037e-05	0.762	0.002144

2. p-value for statistical difference between predictive models for FCE

	Bagging	**K-NN**	**LM**
K-NN	1.18e-13	--	--
LM	0.0005042	1.827e-27	--
NNET	0.08167	1.537e-20	0.08167

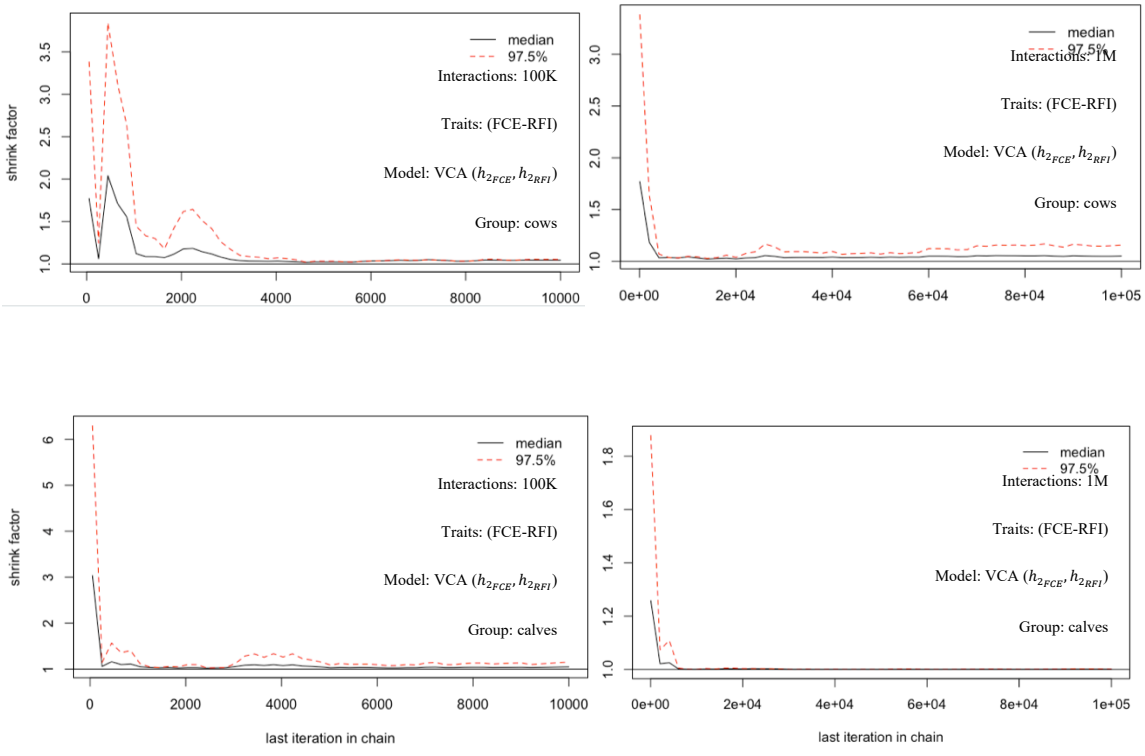
3. p-value for statistical difference between predictive models in RFI

	Bagging	K-NN	LM
K-NN	4.692e-15	--	--
LM	4.692e-15	0.9954	--
NNET	0.1398	2.941e-21	2.941e-21

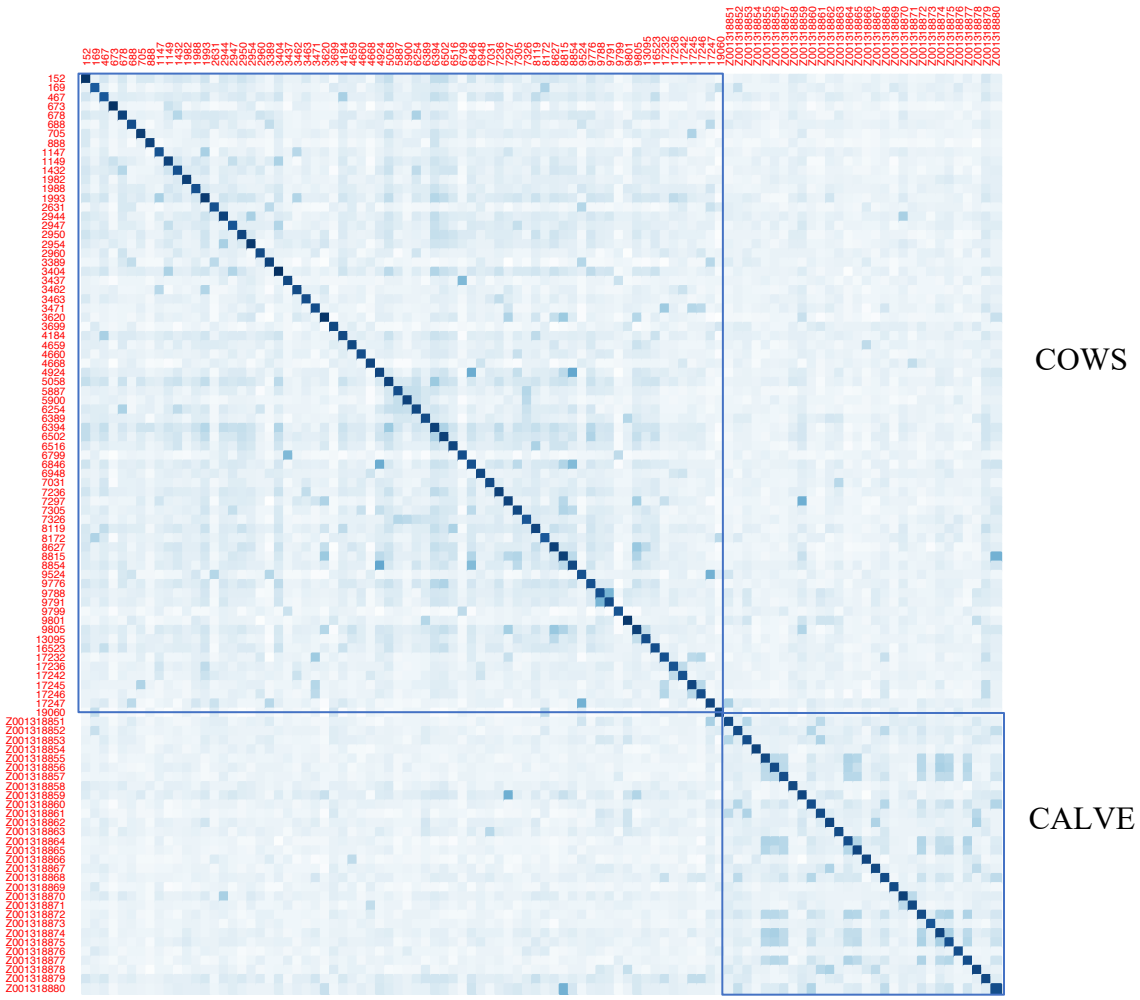
4. p-value for statistical difference between predictive models in ROFC

	Bagging	K-NN	LM
K-NN	7.041e-39	--	--
LM	1.488e-23	1.835e-89	--
NNET	1.995e-24	1.244e-90	0.8003

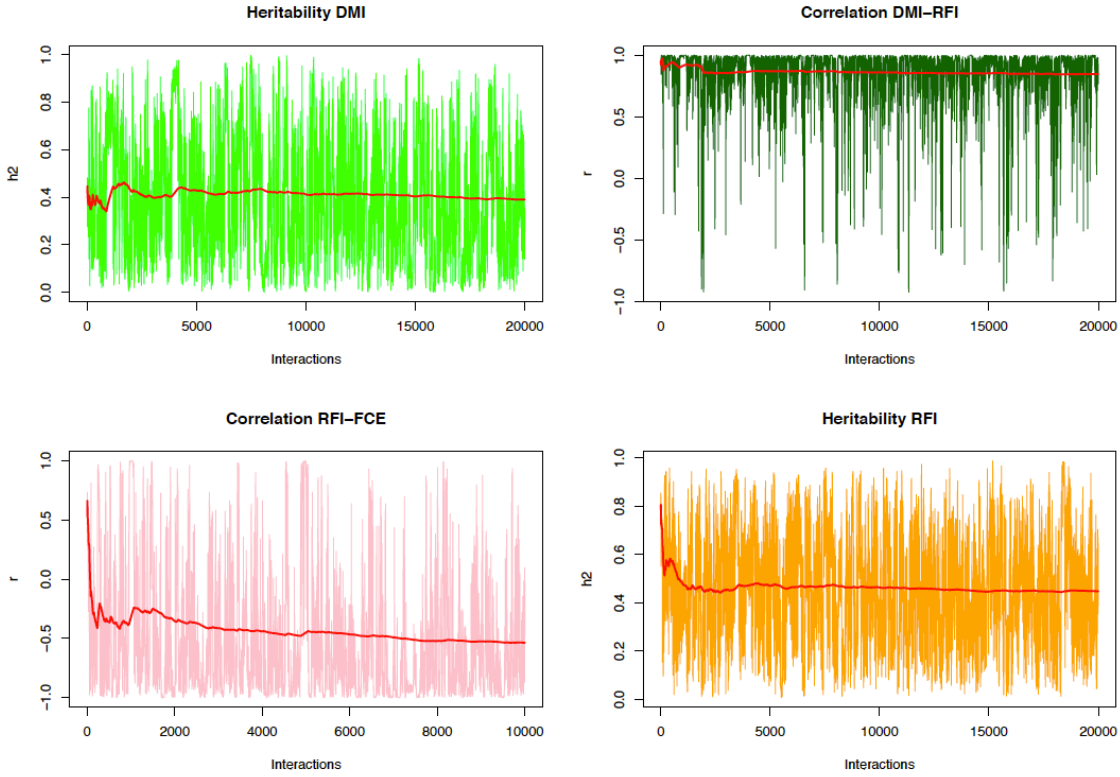
Annex 5. Gelman-Rubin diagnostic for MCMC convergence: Gelman-Rubin diagnostic for convergence was used through the coda package by R. It tests a difference between the variance within some chains and the variance between chains by a value called Potential scale reduction factors. Lower values are desirables. Random traits were selected for the Gelman diagnostic. There was no difference in the diagnosis of convergence between 100 k iterations and 1 M iterations. Convergence chain was also appreciated below.



Annex 6. Heatmap from the genomic relationship matrix (100x100), the colour represents the relation between animals. In red is the ID from the animals. Cows are more related to each other and calves are more related to each other.



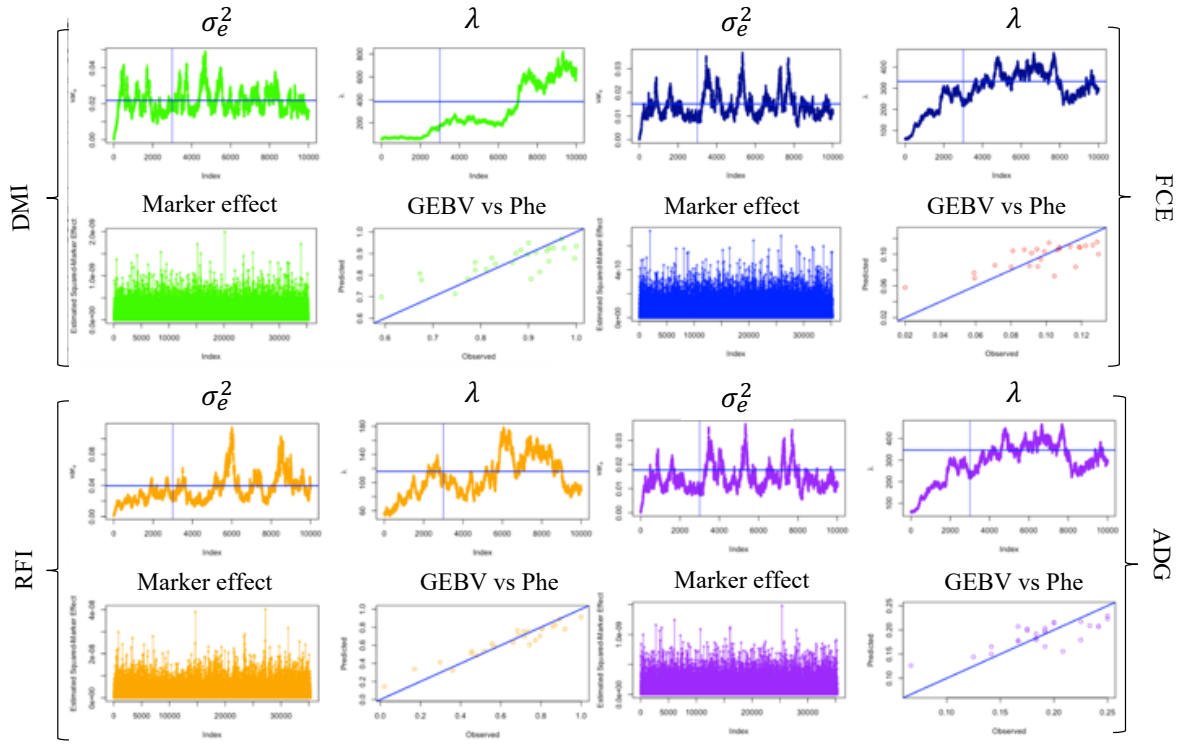
Annex 7. Gibb Samplings interactions and convergence. Random traits were taken to evaluate the convergence in the estimation of heritability and genetic correlations, there was convergence, however the sd was very large.



Annex 8. Priors for lambda in Bayes LASSO associations. Lambda values were selected considering posterior σ_e^2 and λ distributions, and the correlation between GEBV and phenotype (Phe).

	CALVES				COWS			
Lambda	**DMI**	**FCE**	**RFI**	**ADG**	**DMI**	**FCE**	**RFI**	**ROFC**
Type	Random	Random	Random	Random	Random	Random	Random	Random
Shape	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
Rate	1e-5	1e-5	1e-5	1e-5	1e-5	1e-5	1e-5	1e-5
Value	40	40	40	40	20	40	20	40

Annex 9. Posteriori distributions for error variance (σ_e^2) and lambda (λ), marker effects and correlation between predictions values (GEBV) in y axis, and phenotype (Phe) in x axis from Bayes LASSO in calves.



Annex 10. Posteriori distributions for error variance (σ_e^2) and lambda (λ), marker effects and correlation between predictions values (GEBV) in y axis, and phenotype (Phe) in x axis from Bayes LASSO in cows.