



## SNP arrays evaluation as tools in genetic improvement in Corriedale sheep in Uruguay

### Evaluación de paneles de SNP como herramientas en la mejora genética de ovinos Corriedale en Uruguay

### Avaliação de painéis de SNP como ferramentas em melhoramento genético de ovinos Corriedale no Uruguai

Carracelas, B. <sup>1</sup>; Navajas, E. A. <sup>1</sup>; Vera, B. <sup>2</sup>; Ciappesoni, G. <sup>2</sup>

<sup>1</sup>Instituto Nacional de Investigación Agropecuaria (INIA), Unidad de Biotecnología, Las Brujas, Uruguay

<sup>2</sup>Instituto Nacional de Investigación Agropecuaria (INIA), Programa de Producción de Carne y Lana, Las Brujas, Uruguay

#### Editor

Mariana Carriquiry   
Universidad de la República, Facultad de  
Agronomía, Montevideo, Uruguay

Received 09 Feb 2022

Accepted 30 May 2022

Published 19 Aug 2022

#### Correspondence

Beatriz Carracelas,  
bcarracelas@inia.org.uy

#### Abstract

One control strategy for gastrointestinal nematodes (GIN) is genetic selection. This study's objective was to compare eggs per gram of feces (FEC) and fiber diameter (FD) estimated breeding values (EBV) and genomic EBV (GEBV) in Corriedale breed. Analysis included 19547 lambs with data, and 454, 711 and 383 genotypes from 170, 507 and 50K SNP chips, respectively. A univariate animal model was used for EBV and GEBV estimation, which included contemporary group, type of birth and dam age as fixed effects, and age at recording as covariate. Differential weights ( $\alpha$ ) were considered in the genomic relationship matrix (G), and the best fit models were identified using Akaike's Information Criterion (AIC), which were later used for GEBV and accuracies estimation. The use of  $\alpha$  only impacted on low density SNP chips. No differences were observed in mean accuracies for the whole population. However, in the genotyped subgroup accuracies increased by 2% with the 170 SNP chip ( $\alpha=0.25$ ), and by 5% ( $\alpha=0.5$ ) and 14% ( $\alpha=0.75$ ) with the 507 SNP chip. No differences were observed in FD EBV and GEBV mean accuracies. These results show that it is possible to increase GEBV accuracies despite the use of low-density chips.

**Keywords:** accuracy, Corriedale, FEC, GEBV

#### Resumen

Una alternativa para el control de los nematodos gastrointestinales (NGI) es la selección genética. El objetivo de este trabajo fue comparar las precisiones de los valores de cría (EBV) y los EBV genómicos (GEBV) del recuento de huevos por gramo en heces (HPG) y diámetro de fibra (DF) en la raza Corriedale. El análisis incluyó 19547 corderos con datos fenotípicos y 454, 711 y 383 genotipados con paneles o chips de 170, 507 y 50K SNP, respectivamente. Los EBV y GEBV se estimaron con un modelo animal univariado que incluyó los efectos fijos: grupo contemporáneo, tipo de nacimiento y edad de la madre, y edad al registro como covariable. Se consideraron pesos diferenciales ( $\alpha$ ) en la matriz de relaciones genómicas, identificándose los modelos con mejor ajuste con el criterio de información de Akaike (AIC), que fueron utilizados para la estimación de los GEBV y sus precisiones. El uso de  $\alpha$  solo impactó en el ajuste con paneles de baja densidad. No se encontraron diferencias en las precisiones promedio de la población total. En cambio, en el subgrupo de



animales genotipados las precisiones aumentaron 2% con 170 SNP ( $\alpha=0.25$ ), y con 507 SNP 5% ( $\alpha=0.5$ ) y 14% ( $\alpha=0.75$ ). No hubo diferencias en precisiones de los EBV y los GEBV de DF. Los resultados muestran que es posible aumentar las precisiones de los GEBV aun con paneles de baja densidad.

**Palabras clave:** precisión, Corriedale, HPG, GEBV

## Resumo

Uma alternativa para o controle de nematóides gastrointestinais (NGI) é a seleção genética. O objetivo deste trabalho foi comparar as precisões dos valores genéticos estimados (EBV) e dos EBVs genômicos (GEBV) da contagem de ovos por grama de fezes (OPG) e diâmetro de fibra (DF) na raça Corriedale. A análise incluiu 19547 cordeiros com dados e 454, 711 e 383 genotipados de 170, 507 e 50K SNPs, respectivamente. Foram estimados os EBV e GEBV com um modelo animal univariado que incluiu efeitos fixos de grupo contemporâneo, tipo de nascimento e idade da mãe e idade no registro (covariável). Pesos diferenciais ( $\alpha$ ) foram considerados na matriz de relações genômicas, identificando os modelos com melhor ajuste via critério de informação de Akaike (AIC), os quais foram utilizados para estimar o GEBV e suas precisões. O uso de  $\alpha$  somente impactou no ajuste com painéis de baixa densidade. Não foram encontradas diferenças na precisão média da população total. Em contraste, no subgrupo de animais genotipados as precisões aumentaram 2% com 170 SNPs ( $\alpha = 0.25$ ), e com 507 SNPs 5% ( $\alpha = 0.5$ ) e 14% ( $\alpha = 0.75$ ). Não houve diferenças na precisão de EBV e GEBV de DF. Os resultados mostram que é possível aumentar a precisão de GEBVs mesmo que se utilizem painéis de baixa densidade.

**Palavras-chave:** precisão, Corriedale, OPG, GEBV

## 1. Introduction

Uruguay's main production system is based on natural pastures. Because of favorable climatic conditions, sheep can graze all year long, but this has the disadvantage of animals being exposed to gastrointestinal nematodes (GIN). These are the most prevalent parasitic infections in grazing sheep worldwide, which cause important economic losses to the sheep industry, because of the decrease in meat and wool production, as well as the increase in costs associated to anthelmintic control. In Australia, Lane and others<sup>(1)</sup> estimated losses of AUD 436 million per year.

Anthelmintic drugs are used as the main control method, but its frequent and indiscriminate use has favored anthelmintic resistance development. This situation resulted in the onset of alternative control strategies, one of them being the selection of genetically resistant animals. Animal's ability to resist parasitic infections is genetically determined, with variability between breeds as well as between individuals from the same breed<sup>(2)</sup>. Resistance is the ability of an animal to initiate and maintain an immune response to prevent or eliminate a parasitic infection after it is installed, and it is quantified through nematodes egg count per gram of feces (FEC). In Uruguay, since 1994, GIN genetic resistance is included in the Corriedale National Genetic Evaluation ([www.geneticaovina.com.uy](http://www.geneticaovina.com.uy)), using FEC measured in lambs as selection criterion and estimated breeding value (EBV) estimation with the Best Linear Unbiased Predictor methodology (BLUP)<sup>(3)</sup>.

FEC is a moderately heritable trait ( $h^2 \sim 0.3$ ) with a great variability between individuals<sup>(4-6)</sup>.

Nowadays, in most countries, the EBV estimation for selection of genetically superior animals is based on genealogical and phenotypic records, but in the last years, medium-density single nucleotide polymorphism (SNP) arrays have emerged as an additional source of information. Genomic information is a complementary tool in genetic evaluations for low heritability, long generation interval or difficult to measure traits. Genomic selection (GS) is a type of marker assisted selection where a great number of genetic markers (mostly SNP) distributed along the genome are in linkage disequilibrium (LD) with genomic regions associated to quantitative trait loci (QTL)<sup>(7)</sup>. The procedure consists of estimating all SNP effects simultaneously from individuals with phenotypic and genotypic information (reference population), and later using these effects to predict genomic EBV (GEBV) from selection candidates that don't have phenotypic records<sup>(8)</sup>.

Each SNP effect can be estimated using different assumptions about its distribution<sup>(9)</sup>. In genomic BLUP (GBLUP) a normal distribution with equal variances for the markers effects is assumed<sup>(7)(10-11)</sup>. Investigations carried out by Hayes and others<sup>(12)</sup>, VanRaden and others<sup>(13)</sup>, and Cole and others<sup>(14)</sup> have showed that assuming equal variance for each SNP produced little or no loss of accuracy for most traits. In 2009, Misztal and others<sup>(15)</sup> proposed to integrate genomic information in a single step genetic evaluation using single step genomic BLUP

(ssGBLUP), where the additive relationship matrix (A) is combined with the SNP based genomic relationship matrix (G) to create an H matrix<sup>(9)(16-17)</sup>. The main idea is to use all available information (phenotypes, genotypes and pedigree) in a model to predict GEBV for all individuals simultaneously<sup>(18)</sup>.

The use of genomic information in GEBV estimation allows to significantly increase genetic gain through EBV accuracy increase in young animals<sup>(7)</sup>. Nowadays, GS has been implemented in several sheep breeding programs in Australia<sup>(19-20)</sup>, New Zealand<sup>(21)</sup> and France<sup>(22-23)</sup>. Concerning EBV accuracies, Auvray and others<sup>(21)</sup> reported increases between 0.09 and 0.37 when using GBLUP instead of BLUP, for eight traits and four different breeds, with a genotyped population of 13420 animals. Accuracies for milk production trait increased from 0.26 to 0.42 when ssGBLUP method was used<sup>(23)</sup>, and between 0.05 and 0.3 increases for the same trait in six different dairy sheep breeds<sup>(24)</sup>. Regarding FEC trait, Torres and others<sup>(25)</sup> reported EBV accuracies increases between 0.046 and 0.073.

In 2014, Periasamy and others<sup>(26)</sup> identified 170 SNP associated with 76 candidate genes involved in immune response to GIN resistance, and Raschia and others<sup>(27)</sup> found that eight of them were significantly associated with FEC in Corriedale sheep under artificial infection. In turn, INIA [National Agricultural Research Institute] developed a 507 SNP chip with the Affymetrix company that contains 174 SNP related to FEC trait and 258 paternity SNP (FMV\_2\_2011\_1\_6356\_ANII project)<sup>(28)</sup>.

Even though this study is focused on GIN genetic resistance, with FEC as selection criterion, we also included fiber diameter (FD) for comparison reasons since it is a highly heritable trait and EBV accuracies are affected by this parameter<sup>(29)</sup>. This trait is also a relevant selection breeding objective for the Corriedale breed in Uruguay<sup>(30)</sup>.

The aim of this study was to evaluate the contribution of three different SNP arrays in GEBV accuracy increase as compared to traditional genetic evaluation for FEC and FD traits. Due to the small number of genotyped animals, this study was focused on GEBV accuracies.

## 2. Materials and methods

### 2.1 Phenotypic Data

From 2000 to 2019, FEC and FD records were collected from 19547 Corriedale animals belonging to 29 farms (24 stud flocks, 3 experimental units and 2

progeny testing centrals). Genealogical information from 40056 animals was provided by Uruguay's Rural Association (ARU) and the Uruguayan Corriedale Breeders Society, who also provided the performance and management data.

FEC as well as FD traits are routinely registered at the National Genetic Evaluation, but the first one is not mandatory. FEC sampling was performed at 278±69 days mean age according to the protocol used for genetic evaluations<sup>(30)</sup> and FD sampling was made at shearing (364±42 days mean age). Due to FEC's non normal distribution, data was transformed to natural logarithm, Log<sub>e</sub> (FEC+100) as described in Ciappesoni and others<sup>(31)</sup>. Descriptive statistics for both traits are presented in Table 1.

Table 1. Descriptive statistics for fecal egg counts (FEC), Log<sub>e</sub> (FEC+100) and fiber diameter (FD) for N=19547

Trait	Mean	SD	Min	Max
FEC	1310	2158	0	37400
Log <sub>e</sub> (FEC+100)	6.49	1.24	4.61	10.53
FD (μ)	24.98	3.03	16.60	38.80

Fecal samples were collected from the animal's rectum using plastic bags, identified, stored with ice gel packs and taken to the laboratory as described at the INIA N°6 Card<sup>(32)</sup>. Samples were processed at parasitology laboratories at INIA Tacuarembó, INIA Las Brujas or the Uruguayan Wool Secretariat (SUL, by its Spanish acronym), where FEC were assessed using a modified McMaster technique with a sensitivity of 100 eggs per gram of feces<sup>(33)</sup>.

### 2.2 Genotypic Data

Blood samples were collected by jugular's vein puncture using tubes with K<sub>2</sub> EDTA anticoagulant (BD Vacutainer, USA), and afterwards DNA was extracted according to Medrano and others<sup>(34)</sup> protocol with modifications. NanoDrop 8000 spectrophotometer (Thermo Scientific, USA) was used for DNA quantification and purity evaluation. DNA integrity was checked with a 1% agarose gel with 0.5X TBE buffer (Tris-Borate-EDTA, Thermo Scientific, USA) during 25 minutes at 100V. Finally, DNA samples were stored at -80°C until genotyping.

For this study, genotypes from three different SNP chips were used: 454 animals with 170 SNP (International Atomic Energy)<sup>(26)</sup>, 711 animals with 507 SNP (Charrúa Panel, Affymetrix)<sup>(35-36)</sup> and 383 animals with 50K SNP (Illumina Ovine SNP50 Bead-Chip v1 and v2, Affymetrix Oviser Axiom 50K). For

the 50K chip, 33236 SNP in common between Illumina and Affymetrix platforms were used.

For genomic data quality control the PREGSF90 program was used<sup>(37-38)</sup> and consisted in the exclusion of sexual markers, monomorphic, minor allele frequency <0.05 and call rate <90%; and the exclusion of individuals with call rate <90%. SNP and individuals after quality control are presented in Table 2. Regarding SNP in common, only the 507 SNP chip has some markers from the 50K SNP chip (91 SNP).

Table 2. Number of SNP and individuals after quality control. G matrix weights ( $\alpha$ ) and A matrix weights ( $\beta$ ) calculated according to VanRaden<sup>(11)</sup> for 50K, 507 and 170 SNP chips

Chip	SNP	Animals	A	B
50K	29325	375	0.99	0.01
507	373	702	0.75	0.25
170	148	454	0.75	0.25

### 2.3 Animal Welfare

Fecal samples extraction protocol as well as blood extraction protocol were approved by INIA's Animal Ethics Committee (Approval number INIA\_2018.2).

### 2.4 Statistical Analysis

EBVs were estimated based on phenotypic and pedigree records through BLUP methodology. GEBVs were estimated based on phenotypic and pedigree records as well as genotypes through ssGBLUP methodology. Variance components were estimated using the AIREML algorithm (Average Information Restricted Maximum Likelihood)<sup>(39)</sup> and afterwards used as initial values for EBV and GEBV prediction using BLUPF90 family of programs<sup>(40)</sup>.

For breeding values estimation and (co)variance components a univariate animal model was used:

$$y = Xb + Zu + e$$

where  $y$  is the observations vector for each trait ( $\text{Log}_e(\text{FEC}+100)$  or FD);  $X$  is the fixed effects incidence matrix;  $b$  is the vector of fixed effects;  $Z$  is the additive genetic effects incidence matrix;  $u$  is the vector of direct additive genetic effects, and  $e$  is the residual effects vector. Under the infinitesimal model, it is assumed that  $u \sim N(0, A\sigma_u^2)$  with pedigree-based approach, and that  $u \sim N(0, H\sigma_u^2)$  with genomic approach (being  $\sigma_u^2$  the additive genetic variance) and that  $e \sim N(0, I\sigma_e^2)$ . Fixed effects in-

cluded in the model were as follows: 467 contemporary groups (birth year, sex, stud-flock, and management group), type of birth (2 levels: unique or multiple), mother's age (3 levels: 2, 3 and  $\geq 4$  years) and age at recording as covariate (age at FEC and age at shearing).

$H$  is the relationship matrix that combines pedigree and genomic information and was estimated using ssGBLUP. This matrix inverse was calculated according to Aguilar and others<sup>(9)</sup>:

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

where  $A^{-1}$  is the inverse of the pedigree relationship matrix,  $G^{-1}$  is the inverse of the genomic matrix, and  $A_{22}^{-1}$  is the inverse of the pedigree relationship matrix of genotyped animals.

G matrix calculation was computed according to VanRaden's<sup>(11)</sup>:

$$Gw = w * G + (1 - w) * A_{22}$$

where the weighted G matrix ( $Gw$ ) is used with a formula that includes the number of SNPs ( $m$ ):

$$w = \frac{0.05^2}{(0.05^2 + \frac{0.125}{m})}$$

PREGSF90 program<sup>(37)</sup> uses the same formula but with  $\alpha$  and  $\beta$ :

$$G = \alpha G + \beta A_{22}$$

Default values are  $\alpha=0.95$  (G matrix weight) and  $\beta=0.05$  (A matrix weight). These weights are used for G matrix blending, to make it positive definite so it can be inverted<sup>(11)</sup>. Calculated weights for 170, 507 and 50K SNP are presented in Table 2.

The Akaike Information Criterion (AIC)<sup>(41)</sup> was used for model comparison:

$$AIC = 2k - 2\ln(L)$$

where  $2\ln(L)$  is the model's goodness of fit ( $L$  is the maximum likelihood), and  $k$  is a complexity measurement (number of estimated parameters).

AIC values were estimated with the AIREML algorithm. BLUP and ssGBLUP models were compared with different weights assigned to G matrix, and delta AIC was calculated ( $\Delta AIC$ ; difference between two values). According to Burnham & Anderson<sup>(42)</sup>, models with  $\Delta AIC$  less than 2 are equivalent, between 4 and 7 are somewhat different, and  $>10$  are conclusively different. Afterwards, only models with  $\Delta AIC$  higher than 10 were considered, this is to say, models where including SNP information showed a difference compared to BLUP.

To understand the impact of these models on the genetic evaluation, EBV and GEBV accuracies were estimated for the different chips, according to Aguilar and others<sup>(43)</sup>:

$$\text{acc} = \sqrt{1 - \frac{\text{PEV}_i}{(1 + F_i)\sigma_u^2}}$$

where  $\sigma_u^2$  is the additive genetic variance,  $\text{PEV}_i$  is the prediction error variance of animal  $i$  and  $F_i$  is the inbreeding coefficient.

Afterwards, a paired Student's t-test was used to check if these accuracies were statistically significant.

Also, correlation between elements outside the diagonal of  $G$  and  $A_{22}$  matrices were assessed to detect conflicts between pedigree and the genomic matrix. This correlation is expected to be between 0.5 and 0.9, and values higher than 0.9 show that the information between  $G$  and  $A_{22}$  matrices is very similar<sup>(44)</sup>.

### 3. Results

Initially, FEC and FD genetic and residual variances, and heritability estimates were assessed (Table 3), and later used as initial values in breeding values estimation.

Afterwards, AIC for each model was calculated. For FD, only the model with 50K SNP showed  $\Delta\text{AIC}$  values different from BLUP. Furthermore, for FEC this model didn't show any difference, but the models with 170 SNP and 507 SNP did (Table 4). In the case of models with 507 SNP, the ones with 0.25, 0.5 and 0.75  $\alpha$  values and the model with 170 SNP (with  $\alpha=0.25$ ) showed differences for this trait.

Table 3. Additive genetic variance (gV), residual variance (rV) and heritability ( $h^2$ ) estimates for  $\text{Log}_e$  (FEC+100), and fiber diameter (FD) with standard errors

Trait	gV	rV	$h^2$
$\text{Log}_e$ (FEC+100)	0.16±0.02	0.72±0.01	0.18±0.02
FD	1.89±0.10	1.88±0.07	0.50±0.02

Table 4. Akaike Information Criterion (AIC) and delta AIC ( $\Delta\text{AIC}$ ) values for BLUP and ssGBLUP models (with different weights assigned to G matrix) for  $\text{Log}_e$  (FEC+100), and fiber diameter (FD) for 170, 507 and 50K SNP

	$\text{Log}_e$ (FEC+100)		FD	
	AIC	$\Delta\text{AIC}$	AIC	$\Delta\text{AIC}$
BLUP	52489.5		78358.8	
50K SNP $\alpha=0.75$	52484.6	4.9	78337.8	21
50K SNP $\alpha=0.95$	52490.2	-0.7	78336.1	22.7
50K SNP $\alpha=0.99$	52491.7	-2.2	78335.9	22.9
507 SNP $\alpha=0.25$	52473.1	16.4	78357.6	1.2
507 SNP $\alpha=0.50$	52473.5	16	78366.4	-7.6
507 SNP $\alpha=0.75$	52478.6	10.9	78386	-27.2
507 SNP $\alpha=0.95$	52487.4	2.1	78417.8	-59
170 SNP $\alpha=0.25$	52479.4	10.1	78357.3	1.4
170 SNP $\alpha=0.50$	52486.5	3	78369.4	-10.6
170 SNP $\alpha=0.75$	52501.6	-12.1	78396.7	-37.9
170 SNP $\alpha=0.95$	52525.9	-36.4	78445.9	-87.1

To explore which of the AIC selected models optimizes trait selection, mean accuracies for the whole population and for genotyped animals were estimated. BLUP estimations were compared to ssGBLUP (with different weights assigned to G matrix) (Table 5).

Table 5. Mean accuracies estimated with BLUP (EBV Acc) and ssGBLUP (GEBV Acc), and percentage increase for all animals (N=37018) and for genotyped animals with 170 SNP chip ( $\alpha=0.25$ ), with 507 SNP chip (507<sub>1</sub>  $\alpha=0.25$ ; 507<sub>2</sub>  $\alpha=0.50$ ; 507<sub>3</sub>  $\alpha=0.75$ ) or 50K SNP chip (50K<sub>1</sub>  $\alpha=0.75$ ; 50K<sub>2</sub>  $\alpha=0.95$ ; 50K<sub>3</sub>  $\alpha=0.99$ ) for AIC>10 models, for  $\text{Log}_e$  (FEC+100), and fiber diameter (FD). Significant values (\*=P<0.05)

Trait	All				Genotyped			
	EBV Acc	GEBV Acc	Increase (%)	Chip	n	EBV Acc	GEBV Acc	Increase (%)
FEC	0.46	0.46	0	170	407	0.59	0.61	3*
	0.46	0.46	0	507 <sub>1</sub>	639	0.58	0.58	0
	0.46	0.46	0	507 <sub>2</sub>	639	0.58	0.62	7*
	0.46	0.46	0	507 <sub>3</sub>	639	0.58	0.67	16*
FD	0.61	0.61	0	50K <sub>1</sub>	313	0.79	0.79	0
	0.61	0.61	0	50K <sub>2</sub>	313	0.79	0.79	0
	0.61	0.61	0	50K <sub>3</sub>	313	0.79	0.79	0

No differences between EBV and GEBV were found for the whole population, with mean accuracy values of 0.46 and 0.61 for FEC and FD, respectively. It was observed that EBV mean accuracies were higher for the genotyped group compared to the whole population, but this is not surprising since samples that are selected for genotyping are the most informative animals (rams and dams with many offspring) and this impacts directly on the accuracies.

When whole population's EBV and GEBV accuracies were plotted (for models selected with AIC) it was observed that, for FEC, there were animals that lowered their accuracies when the 170 and 507 SNP chips with  $\alpha=0.25$  were used (Figures 1 and 2), but this tendency decreased when the 507 SNP chip with  $\alpha=0.5$  y  $\alpha=0.75$  was used (Figure 2). The 507 SNP chip with  $\alpha=0.75$  showed the highest increases in accuracies. Concerning dams not genotyped with genotyped offspring, higher accuracies were observed when the 507 SNP chip with higher  $\alpha$  was used (Figures 1 and 2).

Figure 1. EBV (Acc) and GEBV (Acc\_170\_0.25 with  $\alpha=0.25$ ) correlation for all animals evaluated for FEC trait ( $r=0.99$ )

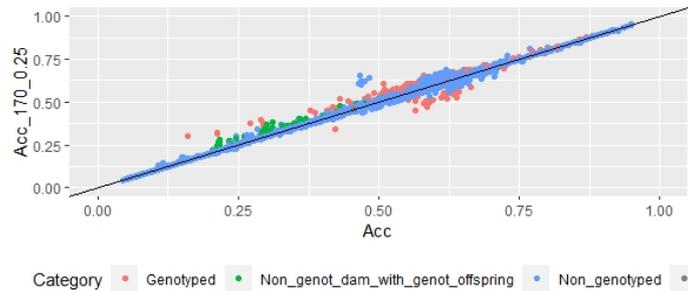
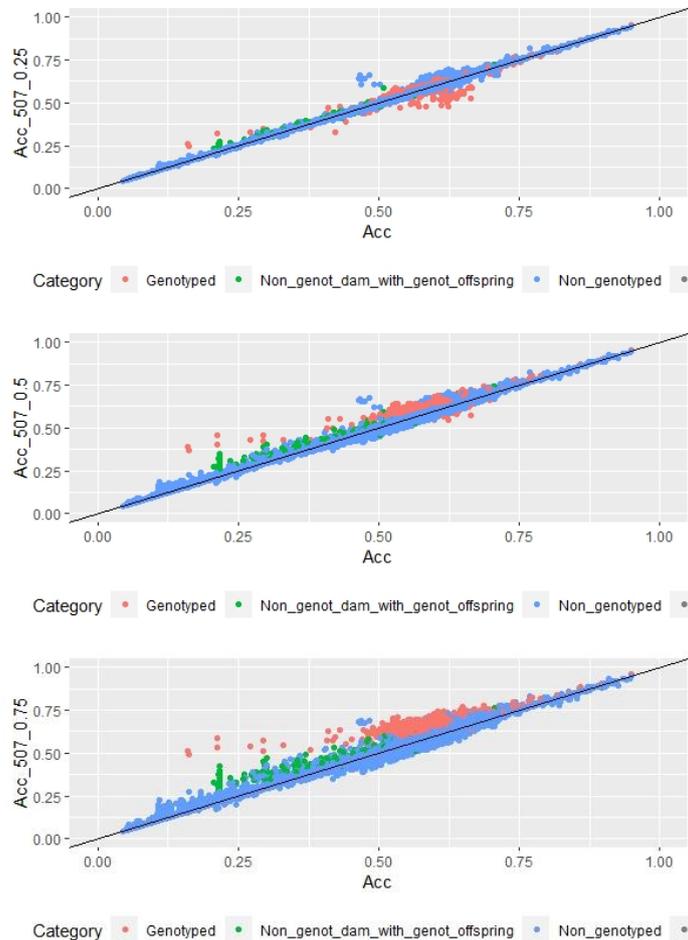


Figure 2. EBV (Acc) and GEBV (Acc\_507\_0.25 with  $\alpha=0.25$ ; Acc\_507\_0.5 with  $\alpha=0.5$  and Acc\_507\_0.75 with  $\alpha=0.75$ ) correlation for all animals evaluated for FEC trait with its correlation coefficients ( $r=0.99$ ,  $r=0.99$ , and  $r=0.99$ , respectively)



Regarding FD, no differences were observed when different  $\alpha$  values were used (Figure 3).

On the other hand, differences in accuracies were observed within the genotyped subset (Table 5). Concerning FEC trait, the 50K SNP chip didn't increase estimated breeding value mean accuracies, and neither did the 507 SNP chip with  $\alpha=0.25$ . By contrast, the 170 SNP chip with  $\alpha=0.25$  and the 507 SNP with  $\alpha=0.5$  and  $\alpha=0.75$  increased GEBV accuracies by 3, 7 and 16%, respectively. These percentages are not comparable between chips since only 3% of animals were genotyped with the three chips.

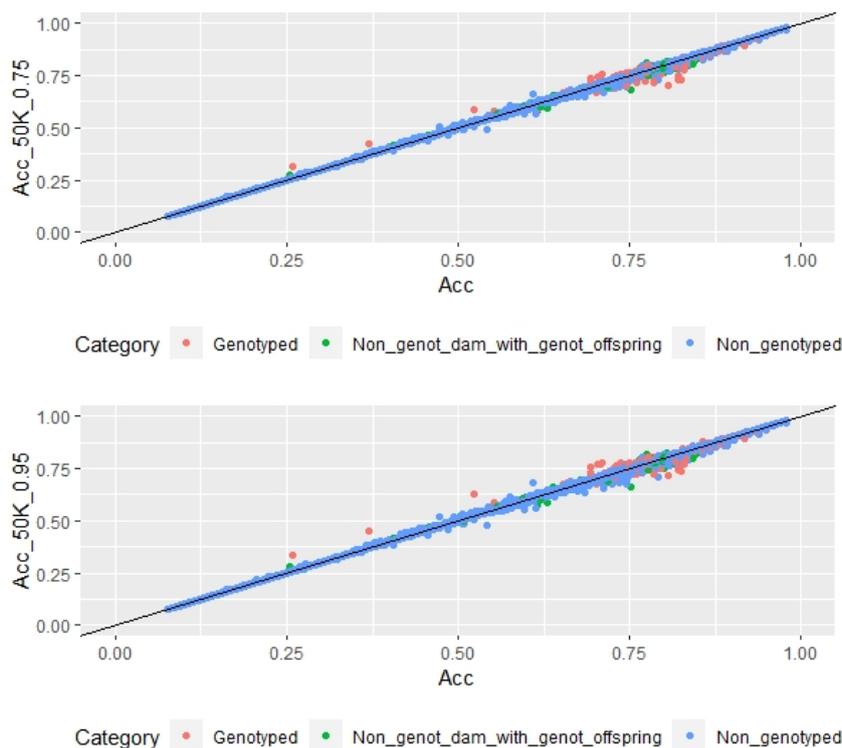
For animals genotyped with both 170 and 507 SNP chips ( $n=305$ ) it was observed that for FEC trait accuracies lowered slightly but were still significant for 170 SNP with  $\alpha=0.25$ , and 507 SNP with  $\alpha=0.5$  and  $\alpha=0.75$ , with 2, 5 and 14% increases, respectively. For 507 SNP with  $\alpha=0.25$  no difference in GEBV accuracy was observed.

Scatter plots for animals genotyped with both 170 and 507 SNP chips for FEC (Figures 4 and 5) showed a strong positive correlation between EBV and GEBV mean accuracies. In addition, it was observed a lower correlation coefficient and a higher accuracy increase when the 507 SNP chip was

used and as  $\alpha$  was increased. Also, in these plots it can be observed that even though a high proportion of animals increased their GEBV accuracies, others with phenotypic data lowered their accuracies when ssGBLUP was used, both with the 170 SNP chip (with  $\alpha=0.25$ ) as well as with the 507 SNP chip (with  $\alpha=0.75$ ). In the case of the 507 SNP chip with  $\alpha=0.75$ , only one animal decreased its GEBV accuracy. It was also observed that animals that increased in higher proportion their accuracies were the animals with lower initial accuracies (animals without phenotypic data and dams whose offspring had phenotypic data). The 507 SNP chip (with  $\alpha=0.75$ ) was the one that yielded higher increases. There was a third group of data that belonged to animals with higher initial accuracies (mostly rams with many offspring that had phenotypic data), that increased their accuracies slightly when the 507 SNP chip (with  $\alpha=0.75$ ) was used, and practically remained without changes when the 170 SNP chip (with  $\alpha=0.25$ ) and the 507 SNP chip (with  $\alpha=0.5$ ) were used.

Regarding G and  $A_{22}$  matrices off-diagonal elements, 0.9, 0.8 and 0.7 values were observed when 170 SNP chip (with  $\alpha=0.25$ ), 507 SNP chip (with  $\alpha=0.5$ ) and 507 SNP chip (with  $\alpha=0.75$ ) were used, respectively.

Figure 3. EBV (Acc) and GEBV (Acc\_50K\_0.75 with  $\alpha=0.75$ , Acc\_50K\_0.95 with  $\alpha=0.95$ , and Acc\_50K\_0.99 with  $\alpha=0.99$ ) correlation for all animals evaluated for FD trait with its correlation coefficients ( $r=0.99$ ,  $r=0.99$  and  $r=0.99$ , respectively)



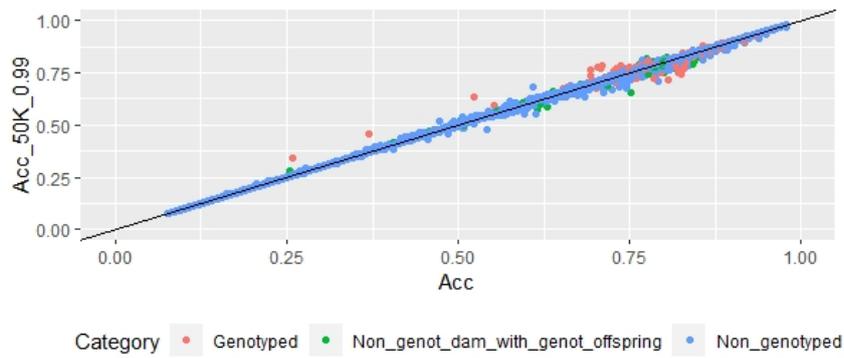


Figure 4. Correlation of EBV (Acc\_FEC) and GEBV accuracies (Acc\_FEC\_170\_0.25 with  $\alpha=0.25$ ) for genotyped animals with the 170 SNP chip, evaluated for FEC trait ( $r=0.93$ )

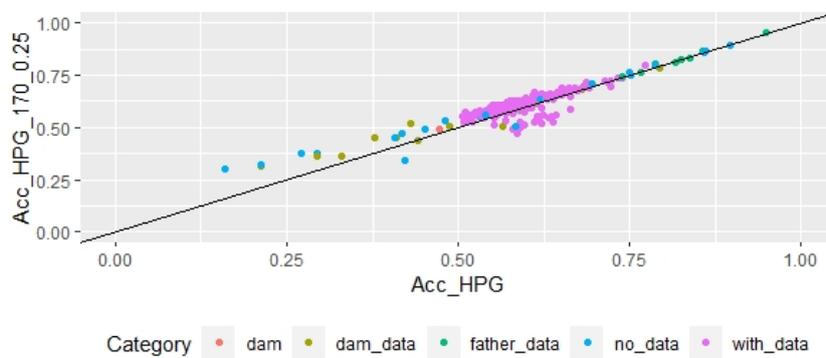
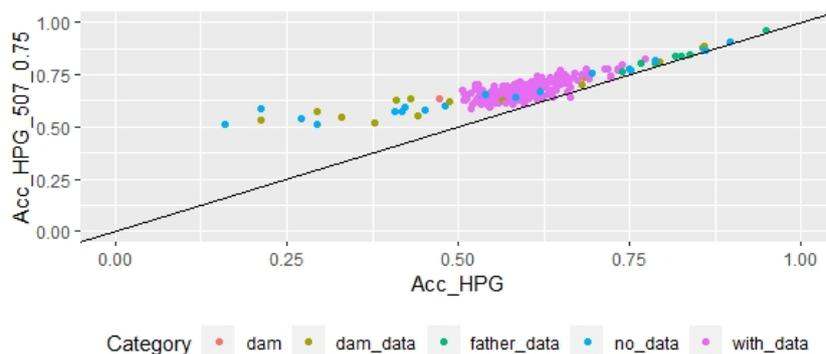
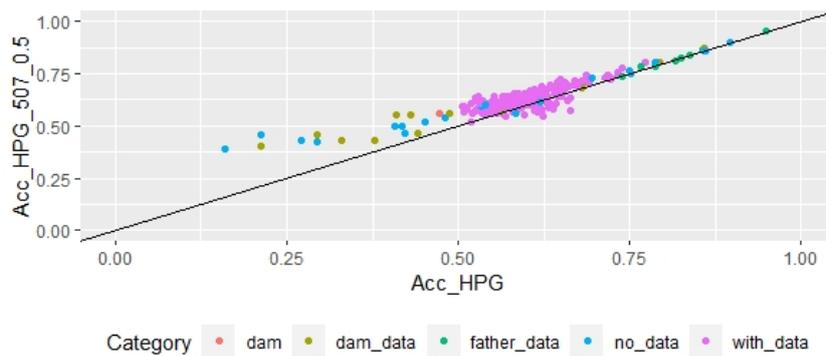


Figure 5. Correlation of EBV (Acc\_FEC) and GEBV accuracies (Acc\_FEC\_507\_0.5 with  $\alpha=0.5$  and Acc\_FEC\_507\_0.75 with  $\alpha=0.75$ ) for genotyped animals with the 507 SNP chip, evaluated for FEC trait with its correlation coefficients ( $r=0.91$  and  $r=0.88$ , respectively)



## 4. Discussion

Molecular contribution to genetic improvement has been widely studied for more than two decades with the objective of increasing the rate of genetic progress. Initial strategies were based on using a few molecular markers associated to the trait of interest (marker assisted selection), but this approach didn't progress due to the polygenic nature of most economically important traits, and therefore, phenotypic effects were too small to be statistically significant<sup>(45)</sup>. Further developments have expanded this concept through the identification of thousands of known SNPs. Commercially available high-density genotyping arrays and different analysis models have made genomic selection implementation possible. One of these models, ssGBLUP, has allowed improvements in GEBV accuracies in sheep production traits compared to BLUP traditional model<sup>(23-25)</sup>.

Since genotyping costs in sheep are quite high relative to the animal's economic value, other alternatives were sought, like the use of low-density SNP chips like the ones that were used in this study.

One way to compare models is cross-validation, where the population is divided in two subsets: training and validation, but in this work, the low number of genotyped animals didn't allow the use of this methodology and instead AIC was used. In 2014, Bernal-Vasquez and others<sup>(46)</sup> compared cross-validation and AIC, and found that both methodologies selected the same models; therefore, AIC could be used as a reliable alternative method. The AIC methodology allowed to select those models that optimized FEC and FD traits compared to BLUP traditional model, that uses only genealogical and phenotypic information. FEC models with 170 SNP chip (with  $\alpha=0.25$ ) and 507 SNP chip (with  $\alpha=0.25$ ,  $\alpha=0.5$  and  $\alpha=0.75$ ) were found to be better than the traditional model. By contrast, using the 50K SNP chip model for FEC showed no difference to traditional model ( $\Delta AIC < 10$ ). For FD, only models with 50K SNP chip with  $\alpha=0.75$ ,  $\alpha=0.95$  and  $\alpha=0.99$  were better than the traditional model.

When GEBV accuracies were estimated for the whole population for those models previously selected with AIC, no differences in mean accuracies were observed with the use of molecular information, probably due to the low percentage of genotyped animals compared to the total evaluated population (between 0.8 and 1.7). Nevertheless, an increase could be observed in dams not genotyped,

but which had genotyped offspring, and in genotyped animals, mostly when the 507 SNP panel with higher  $\alpha$  was used.

For the genotyped subgroup, significant differences in mean accuracies were observed in all models selected with AIC, except the 507 SNP chip model (with  $\alpha=0.25$ ) for FEC trait and all 50K chip models for FD trait. In the case of FD, the 50K SNP chip didn't increase mean accuracies probably due to the low number of genotyped animals (0.8%), and also because FD is a trait with higher heritability compared to FEC ( $h^2=0.50$  versus  $h^2=0.18$ ), so the margin to increase accuracies is much lower. FEC and FD reported heritability estimates for this study agree with previous ones in Corriedale breed in Uruguay<sup>(30)(47-48)</sup>. Heritability is a factor that influences estimated breeding value accuracy: higher trait heritability, higher accuracies.

With the inclusion of the 170 SNP chip in the genetic evaluation a 2% increase in GEBV mean accuracies was observed. Regarding the 507 SNP chip, GEBV mean accuracies increased 5 and 14% when 0.5 and 0.75 G matrix weights were used. In addition, it was observed that animals with lower initial accuracies were the ones that increased their accuracies to a larger extent, and that animals that benefited less were the ones that already had higher accuracies. This study shows that the use of the 507 SNP chip increased GEBV mean accuracy, and those increases were enhanced with higher G matrix weights. These estimates are lower than the ones reported by other authors for GIN resistance; for example, Torres and others<sup>(25)</sup> reported EBV accuracy increases from 0.046 to 0.073 when ssGBLUP was used compared to BLUP, but in that study a 50K SNP chip was used, therefore, with higher SNP density. The 170 and 507 SNP chips are low-density arrays, so they have low genome coverage, and they can't be used for genomic selection where the premise is that a great number of SNP distributed across the genome are in linkage disequilibrium with genomic regions associated to quantitative traits<sup>(7)</sup>. This agrees with a study done in fish populations where prediction accuracies were calculated using different low-density chips, and they found that arrays with less than 1000 SNP show a sharp decrease in accuracies as well as in estimated heritability<sup>(49)</sup>. The only chip utilized here that could be used for GS is the 50K array, but no significant differences were observed in mean GEBV accuracies, probably due to the low number of genotyped animals.

But which could be the reason for the increase in mean accuracies when the 170 and 507 SNP chips were used? EBV accuracy measures the quantity of information used in the prediction of that breeding value (own performance and family performance), and when genomic information is added, the quantity of information increases due to a better capture of family relationships. SNP based pedigree includes information on unrecorded pedigrees and on Mendelian sampling<sup>(12)</sup>, which translate in better accuracies in the estimations. The 507 SNP chip, with  $\alpha=0.75$ , better performance could be related to the higher weight given to the molecular information (G matrix) compared to the pedigree (A matrix). Also, even though the 507 SNP panel is a low-density panel, it is 2.5 times denser than the 170 SNP panel, so theoretically there are more chances for SNPs to be in LD with QTLs related to FEC resistance.

According to Lourenco and others<sup>(44)</sup>, the correlation between elements outside the diagonal of G and  $A_{22}$  matrices is expected to be between 0.5 and 0.9, and values higher than 0.9 indicate that both matrices are very similar, thus a small gain in accuracy is expected. In this study, the 170 SNP chip (with  $\alpha=0.25$ ), the 507 SNP chip (with  $\alpha=0.5$ ) and the 507 SNP chip (with  $\alpha=0.75$ ) show expected values, but the lower value for the 507 SNP chip (with  $\alpha=0.75$ ) indicates higher differences between G and  $A_{22}$  matrices, thus a higher gain in accuracy would be expected.

## 5. Conclusions

It is possible to increase GEBV accuracies for GIN genetic resistance with the use of the 170 SNP identified by Periasamy and others<sup>(26)</sup>, and mainly with the 507 SNP chip with a 0.75 G matrix allocated weight. In the current study increases were observed only on genotyped animals, not in the whole population. To evaluate changes in this population it would be necessary a much greater number of genotyped animals.

Genomic selection in the sheep industry is only possible nowadays in a few developed countries. However, it is expected that more countries incorporate it to their National Genetic Evaluations, mostly in difficult to measure or low heritability traits. This is going to be possible as genotyping prices get cheaper and more animals could be genotyped. This will allow building an adequate training population with enough quantity of genotyped animals with medium-density chips (50K SNP) to be able to apply genomic selection to the National Genetic Evaluations. GS main strength is that it allows increasing

GEBV accuracy in young animals that still do not have their own phenotypic data. Therefore, this study is a first approximation to the incorporation of genomic data about Corriedale's genetic evaluation in Uruguay, that shows that it is possible to increase breeding values mean accuracies even with the use of low-density chips. More research is needed with more genotyped animals and higher density chips to implement genomic selection in this population.

## Acknowledgements

This research has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 772787 (SMARTER), the National Agency for Investigation and Innovation (ANII in Spanish) - María Viñas Fund (FMV\_2\_2011\_1\_6356), the National Agricultural Research Institute (INIA\_CL\_40), and the International Atomic Energy Agency (IAEA RLA5071). In addition, IAEA funded a fellowship training at Animal Production and Health Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture at Seibersdorf, Austria, for Beatriz Carracelas.

## Author contribution statement

BC, EAN, BV and GC conceived and designed the study. BC performed the analysis and wrote the manuscript. BV contributed to data analysis. EAN, BV and GC revised and edited the manuscript. All authors read and approved the final version.

## References

1. Lane J, Jubb T, Shephard R, Webb-Ware J, Fordyce G. Priority list of endemic diseases for the red meat industries [Internet]. North Sydney: MLA; 2015 [cited 2022 May 31]. 282p. Final Report: B.AHE.0010. Available from: <https://bit.ly/3t8UEa1>.
2. Stear MJ, Wakelin D. Genetic resistance to parasitic infection. *Rev Sci Tech*. 1998;1:143-53.
3. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 1975;31:423-47.
4. Safari E, Fogarty NM, Gilmour AR. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. *Livest Prod Sci*. 2005;92:271-89.
5. Bishop SC, Morris CA. Genetics of disease resistance in sheep and goats. *Small Rum Res*. 2007;70:48-59.

6. Keane OM, Hanrahan JP, McRae KM, Good B. An independent validation study of loci associated with nematode resistance in sheep. *Anim Genet.* 2018;49:265-8.
7. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819-29.
8. Goddard ME, Hayes BJ. Genomic selection. *J Anim Breed Genet.* 2007;124:323-30.
9. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 2010;93:743-52.
10. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. *Genetics.* 2007;177:2389-97.
11. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414-23.
12. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci.* 2009;92:433-43.
13. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel S. Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci.* 2009;92:16-24.
14. Cole JB, VanRaden PM, Connell JRO, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR. Distribution and location of genetic effects for dairy traits. *J Dairy Sci.* 2009;92:2931-46.
15. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci.* 2009;92:4648-55.
16. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 2009;92:4656-63.
17. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol [Internet].* 2010 [cited 2022 May 31];42(1):2. doi:10.1186/1297-9686-42-2.
18. Fangmann A. Genomic and conventional evaluations for fertility traits in pigs [doctoral's thesis]. Göttingen (DE): Universität Göttingen, Fakultät für Agrarwissenschaften; 2018. 152p.
19. Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, van der Werf JHJ, Hayes BJ. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim Prod Sci.* 2010;50:1004-10.
20. Daetwyler HD, Swan AA, van der Werf JH, Hayes BJ. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genet Sel Evol.* 2012;44:1-11.
21. Auvray B, McEwan JC, Newman SAN, Lee M, Dodds KG. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. *J Anim Sci.* 2014;92:4375-89.
22. Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM, Barillet F, Robert-Granié C, Manfredi E. Genomic selection in the French Lacaune dairy sheep breed. *J Dairy Sci.* 2012;95:2723-33.
23. Baloche G, Legarra A, Sallé G, Larroque H, Astruc JM, Robert-Granié C, Barillet F. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J Dairy Sci.* 2014;97:1107-16.
24. Legarra A, Baloche G, Barillet F, Astruc JM, Soulas C, Aguerre X, Arrese F, Mintegi L, Lasarte M, Maeztu F, Beltrán de Heredia I, Ugarte E. Within and across breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech and Basco-Béarnaise. *J Dairy Sci.* 2014;97:3200-12.
25. Torres TS, Sena LS, Santos GVD, Figueiredo Filho LAS, Barbosa BL, Júnior AS, Britto FB, Sarmento JLR. Genetic evaluation of sheep for resistance to gastrointestinal nematodes and body size including genomic information. *Anim Biosci.* 2021;34:516-24.
26. Periasamy K, Pichler R, Poli M, Cristel S, Cetrá B, Medus D, Basar M, Thiruvankadan AK, Ramasamy S, Ellahi MB, Mohammed F, Teneva A, Shamsuddin M, Podesta MG, Diallo A. Candidate gene approach for parasite resistance in sheep—variation in immune pathway genes and association with fecal egg count. *PLoS One [Internet].* 2014 [cited 2022 May 31];9(2):e88337. doi:10.1371/journal.pone.0088337.

27. Raschia MA, Donzelli MV, Medus PD, Cetrá BM, Maizon DO, Suarez VH, Pichler R, Periasamy K, Poli MA. Single nucleotide polymorphisms from candidate genes associated with nematode resistance and resilience in Corriedale and Pampinta sheep in Argentina. *Gene* [Internet]. 2021 [cited 2022 May 31];770:145345. doi:10.1016/j.gene.2020.145345.
28. Ciappesoni G, Goldberg V, Navajas EA, Macedo F. SNP Assisted Breeding for Sheep extensive systems: Utopia or Reality. In: International Plant & Animal Genome, Conference PAG XXIV [Internet]. [place unknown]: PAG; 2016 [cited 2022 May 31]. Available from: <https://pag.confex.com/pag/xxiv/meetingapp.cgi/Paper/22397>.
29. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci*. 2009;92:433-43.
30. Castells D. Evaluación de resistencia genética de ovinos Corriedale a los nematodos gastrointestinales en Uruguay: heredabilidad y correlaciones genéticas entre el recuento de huevos de nematodos y características productivas [master's thesis]. Montevideo (UY): Universidad de la República, Facultad de Veterinaria; 2008. 54p.
31. Ciappesoni G, Gimeno D, Ravagnolo O. Genetic relationships between faecal worm egg count and production traits in Merino sheep of Uruguay. In: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production [Internet]. [place unknown]: WCGALP; 2010 [cited 2022 May 31]. 4p. Available from: <http://www.wcgalp.org/system/files/proceedings/2010/genetic-relationships-between-faecal-worm-egg-count-and-production-traits-merino-sheep-uruguay.pdf>.
32. Rodriguez A, Mederos A, Ciappesoni G. Buenas prácticas para recolección y traslado de materia fecal para examen coprológico: determinación de huevos por gramo (HPG) en ovinos y bovinos. Montevideo: INIA; 2009. 2p. (Cartilla; N°6).
33. Whitlock HV. Some modifications of the McMaster helminth egg counting technique and apparatus. *J Coun Sci Ind Res*. 1948;21:177-80.
34. Medrano JF, Aasen E, Sharrow L. DNA extraction from nucleated red blood cells. *Biotechniques*. 1990;8:43.
35. Grasso N. Identificación de polimorfismos de un solo nucleótido asociados a la resistencia a parásitos gastrointestinales en ovinos Corriedale [grade's thesis]. Montevideo (UY): Universidad de la República, Facultad de Agronomía; 2013. 67p.
36. Macedo F. Marcadores SNP para exclusión de paternidad en las razas Corriedale y Merino Australiano del Uruguay [master's thesis]. Montevideo (UY): Universidad de la República, Facultad de Agronomía; 2017. 67p.
37. Aguilar I, Misztal I, Tsuruta S, Legarra A, Wang H. PREGSF90-POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production [Internet]. [place unknown]: WCGALP; 2014 [cited 2022 May 31]. 3p. Available from: <http://www.wcgalp.org/system/files/proceedings/2014/pregsf90-postgsf90-computational-tools-implementation-single-step-genomic-selection-and-genome.pdf>.
38. Misztal I, Tsuruta S, Lourenco DAL, Aguilar I, Legarra A, Vitezica Z. BLUPF90 Family of Programs [Internet]. Georgia: University of Georgia; 2014 [cited 2022 May 31]. Available from: <http://nce.ads.uga.edu/wiki/doku.php>.
39. Thompson R, Brotherstone S, White IMS. Estimation of quantitative genetic parameters. *Philos Trans R Soc Lond B Biol Sci*. 2005;360:1469-77.
40. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production [Internet]. [place unknown]: WCGALP; 2002 [cited 2022 May 31]. 2p. Available from: <http://www.wcgalp.org/system/files/proceedings/2002/blupf90-and-related-programs-bgf90.pdf>.
41. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19:716-23.
42. Burnham KP, Anderson DR. Akaike's information criterion: 1973. In: Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Heidelberg: Springer; 2002. p. 60-6.

43. Aguilar I, Fernandez EN, Blasco A, Ravagnolo O, Legarra A. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J Anim Breed Genet*. 2020;137:356-64.
44. Lourenco D, Legarra A, Tsuruta S, Masuda Y, Aguilar I, Misztal I. Single-step genomic evaluations from theory to practice: using SNP chips and sequence data in BLUPF90. *Genes (Basel)* [Internet]. 2020 [cited 2022 May 22];11(7):790. doi:10.3390/genes11070790.
45. Meuwissen THE, Hayes BJ, Goddard ME. Genomic selection: A paradigm shift in animal breeding. *Anim Front*. 2016;6(1):6-14.
46. Bernal-Vasquez AM, Möhring J, Schmidt M, Schönleben M, Schön CC, Piepho HP. The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics* [Internet]. 2014 [cited 2022 May 31];15(1):646. doi:10.1186/1471-2164-15-646.
47. Ciappesoni G, Goldberg V. Genetic parameters for body weight, worm resistance, packed cell volume and FAMACHA under natural infestation in Corriedale sheep. In: *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production* [Internet]. [place unknown]: WCGALP; 2018 [cited 2022 May 31]. 6p. Available from: <http://www.wcgalp.org/system/files/proceedings/2018/genetic-parameters-body-weight-worm-resistance-packed-cell-volume-and-famachac-under-natural.pdf>.
48. Balconi Marques C, Goldberg V, Ciappesoni G. Genetic parameters for production traits, resistance and resilience to Nematode parasites under different worm burden challenges in Corriedale sheep. *Vet Parasitol* [Internet]. 2020 [cited 2022 May 31];287:109272. doi:10.1016/j.vetpar.2020.109272.
49. Kriaridou C, Tsairidou S, Houston RD, Robledo D. Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. *Front Genet* [Internet]. 2020 [cited 2022 May 31];11:124. doi:10.3389/fgene.2020.00124.