

Building the foundation for the implementation of genomic selection in American Rambouillet sheep

Andre C. Araujo¹
Ronald M. Lewis²
Luiz F. Brito¹

¹Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

²Department of Animal Science, University of Nebraska- Lincoln, Lincoln, Nebraska, USA

February 18, 2023

The effectiveness of genetic selection depends on reliably evaluating the genetic merit of breeding stock for traits that are of economic importance. With advancements in molecular technologies, information collected at the gene-level can now be incorporated into that genetic evaluation. The outcome is more accurate predictions of breeding values, referred to as genomically enhanced estimated breeding values or GEBV, in younger animals allowing for quicker rates of genetic progress. Although GEBV have become routine to breeding programs in many livestock industries worldwide, they are currently only available in one U.S. sheep breed. This study was designed to address that gap by focusing on another key U.S. breed—the American Rambouillet—that is well known for producing a high-quality wool clip in extensive grazing and rangeland systems.

The study was funded by the American Sheep Industry Association Let's Grow Program, the National Sheep Industry Improvement Center, and the American Rambouillet Sheep Breeders Association. We assessed the impact of including genomic information in the Rambouillet genetic evaluation using data collected by the National Sheep Improvement Program (NSIP). The pedigree of 36,297 animals, along with performance records for post-weaning (22,781) and yearling (5,653) body weight, yearling greasy fleece weight (11,542) and fiber diameter (9,586), and the number of lambs born (15,904), were available. Additionally, 741 of these animals had been genotyped, with about 35,000 genomic markers incorporated into our investigations. By including the genomic information, the accuracy of the breeding value estimates for post-weaning and yearling weights were increased by 41% and 62%, respectively. For number of lambs born, the accuracy increased by 37%. Improvement in the reliability of the fiber diameter evaluation was less at 8%, with no gain observed for greasy fleece weight. Such was perhaps to be expected. Wool traits are highly heritable. When coupled with the considerable performance data available, breeding values were already well estimated. Several different approaches for incorporating genomic information were evaluated. The results were similar justifying our use of the computational most efficient approach, known as single-step genomic best linear unbiased prediction, going forward.

Our key conclusion was that by using genomic information along with pedigree and performance data, we obtained as and often much more accurate genetic evaluation of economically important traits in American Rambouillet sheep. Notably, the GEBV were considerably more accurate for number of lambs born, which can be a challenge to evaluate well in younger animals. More genomic information is needed before this technology can be routinely incorporated into the NSIP genetic evaluation of Rambouillet sheep, which is a goal of a project already underway—Sheep GEMS—in partnership with American Rambouillet breeders. Even so, this study has defined the foundation for that next step, highlighting the opportunity genomics offers to enhance the productivity of American Rambouillet sheep.

ORIGINAL ARTICLE

SNP- and haplotype-based single-step genomic predictions for body weight, wool, and reproductive traits in North American Rambouillet sheep

Andre C. Araujo^{1,2}  | Paulo L. S. Carneiro³ | Hinayah R. Oliveira² |
Ronald M. Lewis⁴ | Luiz F. Brito² 

¹Graduate Program in Animal Sciences, State University of Southwestern Bahia, Itapetinga, Bahia, Brazil

²Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

³Department of Biology, State University of Southwestern Bahia, Jequié, Bahia, Brazil

⁴Department of Animal Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

Correspondence

Luiz F. Brito, Department of Animal Sciences, Purdue University, West Lafayette, 47907, IN, USA.
Email: britol@purdue.edu

Funding information

American Rambouillet Sheep Breeders Association; American Sheep Industry Association Let's Grow program; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil); National Sheep Industry Improvement Center (NSIIC-USDA); The State University of Southwest Bahia

Abstract

Rambouillet sheep are commonly raised in extensive grazing systems in the US, mainly for wool and meat production. Genomic evaluations in US sheep breeds, including Rambouillet, are still incipient. Therefore, we aimed to evaluate the feasibility of performing genomic prediction of breeding values for various traits in Rambouillet sheep based on single nucleotide polymorphisms (SNP) or haplotypes (fitted as pseudo-SNP) under a single-step GBLUP approach. A total of 28,834 records for birth weight (BWT), 23,306 for postweaning weight (PWT), 5,832 for yearling weight (YWT), 9,880 for yearling fibre diameter (YFD), 11,872 for yearling greasy fleece weight (YGF), and 15,984 for number of lambs born (NLB) were used in this study. Seven hundred forty-one individuals were genotyped using a moderate (50 K; $n = 677$) or high (600 K; $n = 64$) density SNP panel, in which 32 K SNP in common between the two SNP panels (after genotypic quality control) were used for further analyses. Single-step genomic predictions using SNP (H-BLUP) or haplotypes (HAP-BLUP) from blocks with different linkage disequilibrium (LD) thresholds (0.15, 0.35, 0.50, 0.65, and 0.80) were evaluated. We also considered different blending parameters when constructing the genomic relationship matrix used to predict the genomic-enhanced estimated breeding values (GEBV), with alpha equal to 0.95 or 0.50. The GEBV were compared to the estimated breeding values (EBV) obtained from traditional pedigree-based evaluations (A-BLUP). The mean theoretical accuracy ranged from 0.499 (A-BLUP for PWT) to 0.795 (HAP-BLUP using haplotypes from blocks with LD threshold of 0.35 and alpha equal to 0.95 for YFD). The prediction accuracies ranged from 0.143 (A-BLUP for PWT) to 0.330 (A-BLUP for YGF) while the prediction bias ranged from -0.104 (H-BLUP for PWT) to 0.087 (HAP-BLUP using haplotypes from blocks with LD threshold of 0.15 and alpha equal to 0.95 for YGF). The GEBV dispersion ranged from 0.428 (A-BLUP for PWT) to 1.035 (A-BLUP for

Ronald M. Lewis and Luiz F. Brito share the last authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Animal Breeding and Genetics* published by John Wiley & Sons Ltd.

YGFW). Similar results were observed for H-BLUP or HAP-BLUP, independently of the LD threshold to create the haplotypes, alpha value, or trait analysed. Using genomic information (fitting individual SNP or haplotypes) provided similar or higher prediction and theoretical accuracies and reduced the dispersion of the GEBV for body weight, wool, and reproductive traits in Rambouillet sheep. However, there were no clear improvements in the prediction bias when compared to pedigree-based predictions. The next step will be to enlarge the training populations for this breed to increase the benefits of genomic predictions.

KEYWORDS

Best Linear Unbiased Predictions, genomic-enhanced estimated breeding values, haplotype prediction, linkage disequilibrium, small ruminants, single-step genomic BLUP

1 | INTRODUCTION

The global demand for products from small ruminants is increasing. Further scientific innovation, with its greater application through increased education and training, is needed to meet this demand (Mazinani & Rude, 2020). Rambouillet is a common sheep breed in the US originating from Spanish Merinos that were first moved to a farm in Rambouillet, France, late in the eighteenth century (American Rambouillet Sheep Breeders Association). Known for producing high-quality meat and wool (Thorne et al., 2022), Rambouillet sheep are commonly raised in extensive grazing systems under a wide range of climatic conditions. Rambouillet sheep produces heavy fleeces with fine fibre diameter (Thorne et al., 2021), which are suitable for the wool market. Yet, as a dual-purpose breed, body weight and reproductive traits are also of economic importance in this breed (Thorne et al., 2021). Estimated breeding values (EBV) for these sets of traits have been generated and shared with US sheep producers through the National Sheep Improvement Program (NSIP; Notter, 1998). However, genomic evaluations are not yet available for the Rambouillet breed in the US.

With the availability of large-scale genomic information, the pedigree relationship matrix (**A**), originally used to obtain EBV from Best Linear Unbiased Predictions (BLUP), can be replaced or combined with the genomic relationship matrix (**G**) to predict genomic-enhanced breeding values (GEBV; Aguilar et al., 2010). The GEBV can be more accurate than EBV especially for young animals (not yet recorded for the traits of interest) and for lowly heritable and sex-limited traits (Meuwissen et al., 2001). Furthermore, GEBV can provide advantages for the evaluation of difficult- or expensive-to-measure traits (Brito et al., 2020; Thorne et al., 2021).

The single-step genomic BLUP (ssGBLUP; Christensen & Lund, 2010; Legarra et al., 2009) is a method that

simultaneously includes both genotyped and non-genotyped individuals in the analysis to obtain GEBV for all individuals by combining the genomic and pedigree information. The ssGBLUP is more compatible with current breeding programs (where not all breeding individuals are genotyped) and provides similar or better results than other methods (Guarini et al., 2018; Legarra et al., 2014). However, an important consideration when implementing the ssGBLUP method is how to weight the genomic and pedigree information (McMillan & Swan, 2017; Meyer et al., 2018). This conundrum arises because as **G** is based on the relationships at the genomic marker level, it can be difficult to invert, may not be on the same scale as the **A**, and may not account for residual polygenic effects (Meyer et al., 2018).

To facilitate the inversion process, and to account for residual polygenic effects, two parameters, α and β (with $\alpha = 0, \dots, 1$ and $\beta = 1 - \alpha$), are commonly used to include a proportion of **A** in the **G** that is used in the genomic evaluation (Meyer et al., 2018). Values between 0.95 and 0.99 are common choices to weight **G** (McMillan & Swan, 2017). However, some authors (Gao et al., 2012; McMillan & Swan, 2017) showed that different α can affect the accuracy and bias of the single-step genomic predictions. McMillan and Swan (2017) used $\alpha = 0.50$ to place equal emphasis on the pedigree and genomic relationships in Terminal-Sire sheep breeds, due to less dispersion and similar accuracies of the GEBV compared to the use of higher α values. Therefore, defining the appropriate value for these parameters is important as they may differ even for different traits in the same population (Gao et al., 2012).

The **G** matrix used in the ssGBLUP can also be computed based on different methods. Fitting single nucleotide polymorphisms (SNP) has been the standard method used in genomic analyses. However, haplotypes can also be used for both genomic prediction (Araujo et al., 2021;

Feitosa et al., 2020; Teissier et al., 2020) and genome-wide association (Araujo et al., 2022; Bovo et al., 2021; Feitosa et al., 2021) studies. Haplotypes are the alleles from a set of adjacent loci (sizable regions called haplotype blocks or haploblocks) expected to be inherited together due to lower recombination (Gabriel et al., 2002). Haplotypes are also expected to be in higher linkage disequilibrium (LD) with the quantitative trait loci (QTL) than the single SNP (Calus et al., 2008) and capture epistatic effects (Hess et al., 2017; Jiang et al., 2018), which could result in higher accuracies and lower bias in the genomic predictions (Araujo et al., 2021; Calus et al., 2008).

Previous studies using haplotypes in genomic predictions in livestock have shown a varying improvement in the accuracies of genomic prediction, ranging from no to small (Mucha et al., 2019) up to 22% (Teissier et al., 2020) increases in GEBV accuracy compared to SNPs. The potential reasons for these inconsistencies are multifaceted. The first one is the genetic background of the populations. Most studies evaluating genomic predictions based on haplotypes were performed in cattle and purebred populations (e.g., Cuyabano et al., 2014, 2015; Feitosa et al., 2020; Mucha et al., 2019; Xu et al., 2020), which have relatively low effective population size (N_e). Furthermore, in these studies, the factors that might contribute to better haplotype-based predictions (e.g., epistasis) were not fully explored. Second, the method used to create the haplotypes can influence the results (Araujo et al., 2021; Hess et al., 2017). Several methods have been investigated including genomic windows with different numbers of SNP (usually in kb) and LD based, for instance, on different threshold levels (Teissier et al., 2020). The coding of the haplotype alleles and how to create the genomic relationship matrices including haplotypes also show differences among studies. Haplotypes tend to be multiallelic and can be either fitted as multiallelic markers in a Genomic Restricted Maximum Likelihood-BLUP (GREML-GBLUP) approach (Da, 2015; Prakapenka et al., 2020) or as biallelic pseudo-SNP (ps-SNP) defined based on the haplotype alleles in a ssGBLUP approach (Teissier et al., 2020). Further exploration of the potential value of using haplotype information in genomic prediction is still needed, particularly in small ruminants.

The implementation of genomic selection is still incipient in small ruminants, mainly due to the relatively high cost of genotyping per animal (Mrode et al., 2018), smaller herd sizes, and lower adoption of reproductive technologies such as artificial insemination (in comparison to cattle). Araujo et al. (2021) hypothesized that fitting haplotypes in genomic predictions could outperform the use of SNP in populations with high N_e because it would better capture the complex interactions within haploblocks. However, these authors did not simulate epistasis

and recommended new studies in real populations. Sheep is a specie in which moderate to high N_e are common in commercial populations (Brito, Mcewan, et al., 2017; Kijas et al., 2012), with predictions of GEBV based on haplotypes still scarce (Araujo et al., 2021). Therefore, we aimed to evaluate the accuracy, bias, dispersion, and theoretical accuracy (TA) of GEBV using ssGBLUP fitting SNP or haplotypes (as ps-SNP) for body weight, wool, and reproductive traits in Rambouillet sheep and make comparisons with traditional pedigree-based genetic evaluations. We also evaluated the effect of constructing the haplotypes with different LD thresholds and using alternative α values when forming the **G** matrix. Finally, recommendations for future steps for the implementation of genomic evaluations in Rambouillet sheep were provided.

2 | MATERIAL AND METHODS

No ethical review and approval were needed for this study because all the datasets used were provided by commercial breeding operations.

2.1 | Phenotypic and pedigree data

The phenotypic and pedigree datasets were provided by the NSIP, which included three body weight traits [birth weight (BWT), postweaning weight (PWT), and yearling weight (YWT)], two wool traits [yearling fibre diameter (YFD) and yearling greasy fleece weight (YGFW)] and one reproductive trait [number of lambs born (NLB)] as described in Table 1. The BWT trait represents the lamb weight recorded within 24 h after birth, while PWT and YWT are the body weights recorded at five to 10 (151–304 days) and 10 to 14 (305–428 days) months of age, respectively. The wool traits were measured at yearling age (10–14 months). The pedigree dataset had 36,297 individuals born from 1985 to 2021, spanning up to 15 generations from animals with phenotypic records and with an average (standard deviation) Pedigree Completeness Index (PCI) of 0.57 (0.34). The number of generations traced back and PCI were calculated using the “optiSel” R package (Wellmann, 2019).

The phenotypic datasets used for the genetic and genomic prediction analyses for the body weight and wool traits were processed previously by the NSIP, which provided preadjusted phenotypes (<http://nsip.org/wp-content/uploads/2015/04/Lambplan-TC-Report-Notter.pdf>). Briefly, the preadjustment considered birth and rearing type (fixed levels as a multiplicative adjustment), and age of dam (measured in days) at recording (quadratic and linear regressions). No preadjustments were done for NLB.

TABLE 1 Description of the datasets used for the genetic and genomic predictions of birth weight (BWT), postweaning body weight (PWT), yearling body weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep

Dataset	Variable ^a	Trait					
		BWT (kg)	PWT (kg)	YWT (kg)	YFD (μm)	YGFW (kg)	NLB (count)
Complete	Average	4.86	35.36	59.99	18.87	3.04	1.71
	SD	1.02	7.78	15.28	1.67	0.81	0.59
	Individuals (<i>n</i>)	28,317	22,781	5653	9586	11,542	6846
	Records (<i>n</i>)	28,317	22,781	5653	9586	11,542	15,904
	CG (<i>n</i>)	427	461	110	195	210	445
	Genotypes (<i>n</i>)	587	632	442	529	502	242
Partial	Average	4.77	35.01	58.03	18.92	2.99	1.79
	SD	0.98	7.12	13.54	1.57	0.72	0.59
	Individuals (<i>n</i>)	22,115	17,118	3657	6808	9055	5407
	Records (<i>n</i>)	22,115	17,118	3657	6808	9055	13,790
	CG (<i>n</i>)	404	402	101	175	198	411
	Genotypes (<i>n</i>)	469	456	341	426	402	138
	Focal (<i>n</i>)	118	176	101	103	100	104

^a Standard deviation (SD); number of phenotyped individuals (Individuals), records (Records), contemporary groups (CG), and genotypes (Genotypes) included in the whole and partial datasets (after quality control); and number of focal (also known as validation) individuals (Focal). All genotyped and focal animals had own phenotypes or progeny with phenotypes. The complete data set contained all corrected phenotypes after quality control, and the partial data set was a subset of the complete data truncated by the birth year of the focal individuals (young selection candidates used to compare the methods evaluated).

The data underwent quality control (QC) with observations deviating more than three standard deviations from the mean removed from further analyses.

Contemporary groups (CG) were created by concatenating flock, year, season, management group, sex, recording date, and 70-day age groups to split lambing (birth) dates into 70-day consecutive periods for PWT, YWT, YFD, and YGFW. For BWT, the CG included all the effects previously mentioned for body weight and wool traits, excluding recording date; 35- rather than 70-day age group were also used. The CG for the NLB were created considering ewe's flock, birth year, season, management group, and parturition number (e.g., ewe's first, second, or third lambing). The preadjusted phenotypes for body weight and wool traits and the NLB phenotypes were then adjusted for the CG effect, so that the phenotypes analysed—henceforth referred to as corrected phenotypes—accounted for all systematic environmental effects considered in the NSIP genetic evaluations. The CG for all analysed traits were treated as fixed effects and, as a final QC step, CG with less than three animals and with no phenotypic variability within CG were removed.

2.2 | Genotypic data

Samples from 741 animals were genotyped using the GeneSeek Genomic Profiler Ovine 50K array (Neogen

Corporation) (52,260 SNP; 677 animals) and BovineHD BeadChips (Illumina Inc.) (606,006 SNP; 64 animals) SNP panels by Neogen (GeneSeek, A Neogen Company). These individuals were chosen to be genotyped based on pedigree-based relatedness (genetic connectedness) aiming to capture as much genetic diversity as possible in animals with DNA samples, coming from nine representative NSIP Rambouillet flocks. Connectedness values were derived from prediction error co-variances among EBV (Kuehn et al., 2007, 2008; Lewis et al., 2005) for preadjusted PWT. In these analyses, an animal model was fitted with CG as fixed, and additive and residual effects as random. As an additional criterion, animals with phenotypic information, either on themselves or on their progeny, for most traits analysed were prioritized for genotyping.

Approximately 35K (35,105) autosomal SNP were in common between the two panels. The QC for the genotypic data was done using the PLINK 1.9 software (Purcell et al., 2007), with SNP markers with MAF < 0.05, call rate < 0.90, extreme departure from Hardy–Weinberg equilibrium ($P < 10^{-8}$), located on non-autosomal chromosomes, or a duplicate on the array removed. Samples with call rate < 0.90 were also removed. A total of 32,584 SNP and 722 samples remained for further analyses. The average distance between adjacent markers was 75.433 kb, with standard deviation, minimum, median, and maximum equal to 67.275 kb, 0.002 kb, 59.492 kb, and 2307.471 kb, respectively, in the 32K panel used for the analyses.

2.3 | Haplotype construction

The SNP genotypes for all samples were phased using the FIMPUTE v.3.0 software (Sargolzaei et al., 2014) to infer the parental inheritance (i.e., which allele came from which parent), before creating the haplotype blocks. LD haploblocks were constructed using the r^2 metric (Hill & Robertson, 1968) with the thresholds of 0.15, 0.35, 0.50, 0.65, and 0.80 and based on the Big-LD approach (Kim et al., 2018). The “GPART” package (Kim et al., 2019) implemented in R (R Core Team, 2020) was used to build the haploblocks.

2.4 | Genetic evaluation

2.4.1 | Pedigree-based predictions

Three linear mixed models for the pedigree-based BLUP (A-BLUP) were used in this study, which are defined as follow:

$$\mathbf{y}_1 = \mathbf{1}'\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

$$\mathbf{y}_2 = \mathbf{1}'\mu + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e}, \quad (2)$$

$$\mathbf{y}_3 = \mathbf{1}'\mu + \mathbf{Z}\mathbf{u} + \mathbf{Z}_2\mathbf{m} + \mathbf{S}\mathbf{q} + \mathbf{e}, \quad (3)$$

where the model (1) is an additive genetic model with \mathbf{y}_1 representing a vector of single corrected phenotypic records, μ is the overall mean, \mathbf{u} is the random direct additive genetic effect, and \mathbf{e} is the random residual. The model (2) is a repeatability model, in which \mathbf{p} is the random permanent environment effect, \mathbf{y}_2 contains the repeated corrected phenotypic records, and the other vectors are the same as in model (1). The model (3) also includes the random maternal additive genetic and maternal permanent environment effects, \mathbf{m} and \mathbf{q} , respectively. The $\mathbf{1}'$ is a vector of ones used to calculate the overall mean and \mathbf{Z} , \mathbf{W} , \mathbf{Z}_2 , and \mathbf{S} are the incidence matrices that relates the corrected phenotypic records to the random direct additive, permanent environment, maternal additive, and maternal permanent environment effects, respectively. The random effects for the above models were assumed to be normally distributed with (co)variance structures as follows:

$$\text{Model (1): Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 \\ 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}, \quad (4)$$

$$\text{Model (2): Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_p^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}, \quad (5)$$

$$\text{Model (3): Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{m} \\ \mathbf{q} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 & 0 \\ 0 & \mathbf{A}\sigma_m^2 & 0 & 0 \\ 0 & 0 & \mathbf{I}\sigma_q^2 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}, \quad (6)$$

where σ_u^2 , σ_p^2 , σ_m^2 , σ_q^2 , and σ_e^2 are the additive genetic, permanent environment, maternal genetic, maternal permanent environment, and residual variances, respectively, and model (1) was used to make the EBV prediction for the YFD, model (2) for NLB, and model (3) for BWT, PWT, YWT, and YGFW.

The BLUPF90 software (Miszta et al., 2018) was used to predict the EBV assuming the variance components were known (Table 2). To be consistent with the national genetic evaluation underway in Rambouillet sheep, the models fitted and the variance components used to predict the EBV were provided by NSIP.

2.4.2 | Single-step genomic BLUP using SNP

The corrected phenotypes, models, and variance components used to predict the GEBV under the single-step genomic BLUP using SNP (H-BLUP) approach were similar to the ones used in A-BLUP, except for the inclusion of genomic relationships from the genotyped individuals. In the assumptions of the H-BLUP, the \mathbf{y} vector had corrected phenotypes for genotyped and non-genotyped animals and $\mathbf{u} \sim N(0, \mathbf{H}\sigma_u^2)$. \mathbf{H} is a relationship matrix that combines the pedigree and the genomic relationship matrices (Legarra et al., 2009), with its inverse computed as follows (Aguilar et al., 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (7)$$

where \mathbf{A}^{-1} is the inverse of the pedigree relationship matrix, \mathbf{A}_{22} and \mathbf{A}_{22}^{-1} are the pedigree relationship matrix for the genotyped animals and its inverse, respectively, and \mathbf{G} is the genomic relationship matrix calculated as proposed by Vanraden (2008):

$$\mathbf{G} = \frac{\mathbf{M}\mathbf{M}'}{2 \sum p_i(1 - p_i)}, \quad (8)$$

where \mathbf{M} has the dimension of n genotyped animals by m SNP markers (coded as 0, 1, or 2 for the absence, presence of one copy, or presence of two copies of the reference allele, respectively) and is centred based on twice of the allelic frequencies (p_i ; $1 - p_i$). The PREGSF90 software

TABLE 2 Variance components and genetic parameters used to predict the estimated breeding values for birth weight (BWT), postweaning body weight (PWT), yearling body weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep

Parameter ^a	BWT	PWT	YWT	YFD	YGFW	NLB
σ_u^2	0.085	3.211	15.402	1.311	0.122	0.025
σ_p^2	–	–	–	–	–	0.009
σ_m^2	0.091	1.926	1.777	–	0.013	–
σ_q^2	0.061	1.926	1.777	–	0.013	–
σ_e^2	0.372	25.046	40.283	0.989	0.181	0.250
σ_p^2	0.610	32.110	59.240	2.300	0.330	0.284
h^2	0.140	0.100	0.260	0.570	0.370	0.090
p^2	–	–	–	–	–	0.030
h_m^2	0.150	0.060	0.030	–	0.040	–
c^2	0.100	0.060	0.030	–	0.040	–

^a σ_u^2 = additive genetic variance, σ_p^2 = permanent environment variance associated with repeated records, σ_m^2 = maternal additive genetic variance, σ_q^2 = maternal permanent environment variance, σ_e^2 = residual variance, σ_p^2 = phenotypic variance, h^2 = heritability for the direct additive genetic effect, p^2 = repeatability, h_m^2 = heritability for the maternal additive genetic effect, c^2 = fraction of the phenotypic variance explained by the maternal permanent environment effect.

(Misztal et al., 2018) was used to create the \mathbf{H}^{-1} matrix including the pedigree information, with τ and ω parameters assumed as the default values (1.0). Different values of α and β were used to evaluate the impact of increasing the proportion of \mathbf{A}_{22} on \mathbf{G} , with $\alpha = 0.95$ and 0.50 , and thereby $\beta = 0.05$ and 0.50 , respectively, in which the former is the default value in the PREGSF90 software and the later was suggested as the choice in Terminal Sire sheep populations (McMillan & Swan, 2017).

2.4.3 | Single-step genomic BLUP using haplotypes

The model and assumptions used in the single-step genomic BLUP including haplotypes (HAP-BLUP) approach were similar to those described for H-BLUP. However, the \mathbf{G} used in the $(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1}$ component was constructed using non-LD-clustered SNP (NCSNP) and ps-SNP. A ps-SNP corresponds to one of the unique haplotype alleles present within a haploblock, coded as 0, 1, or 2 to account for the number of copies of the reference haplotype allele, similar to Teissier et al. (2020) and Araujo et al. (2021). As a haploblock can be multi-allelic, several ps-SNP can be created from a haploblock. The ps-SNP were subjected to the same QC criteria as the SNP before their use for genomic prediction. The number of NCSNP plus ps-SNP before QC ranged between 33,922 and 44,695 with the LD thresholds of 0.80 and 0.15, respectively, while the number of NCSNP plus ps-SNP after QC (markers used in the haplotype predictions) ranged from 32,649 to 39,787 with the same LD thresholds (Appendix S1: Table S1). All the scenarios regarding the different combinations of α and β parameters described for H-GBLUP were also tested in the HAP-BLUP method.

2.5 | Comparing genetic and genomic predictions

The whole (complete) and partial datasets used to compare the genetic and genomic predictions (Legarra & Reverter, 2018) for each trait (Table 1) were defined separately based on time thresholds considering the birth date of the genotyped animals as the reference. The whole datasets included all corrected phenotypic records and genotyped individuals with corrected phenotypes on itself or on its progeny. As the number of genotyped individuals was small, the division into partial datasets considered the following two criteria: (1) at least 100 genotyped individuals with average EBV accuracy higher than 0.50 as focal individuals (selection candidates with masked corrected phenotypes on itself and on its progeny) were kept; and (2) at least 20% of the genotyped individuals as focal individuals were kept. Initially, as the genotyped animals were born between 2009 and 2017, all individuals that were born in 2017 were elected to be focal individuals. However, with this approach it was not possible to keep at least 100 genotyped individuals in the focal set for all traits. Ultimately, different dates were used to divide the whole data in the partial datasets (April 21, 2016 for BWT and PWT, April 13, 2016 for YWT, April 26, 2016 for YFD, April 20, 2016 for YGDW, and 04/06/2016 for NLB).

The number of flocks included in the whole dataset for BWT, PWT, YWT, YFD, YGFW, and NLB were 15, 23, 17, 19, 19, 15, respectively, and in the partial datasets were 13, 13, 11, 12, 12, 12, respectively. We used the variance of the difference in unit effects (VED) to evaluate the connectedness between the genotyped samples considered as training and validation (focal) in the whole and partial datasets. The VED is based on the variance of unit effects,

which approximates the mean prediction error variance (PEV) over unities (Kennedy & Trus, 1993), and, as in the PEV, the lower the value the more connected are the units under consideration. The “GCA” R package (Yu & Morota, 2021) was used to calculate the VED, in which the choice for the VED statistic was done due to calculating the connectedness in a single-step and being less computationally expensive than PEV of the difference. The VED for the individuals used to evaluate the genomic predictions for BWT, PWT, YWT, YFD, YGFW, and NLB were 0.046, 0.059, 0.031, 0.008, 0.017, and 0.282, respectively, suggesting sufficient connectedness as the values were low, especially for BWT, PWT, YWT, YFD, and YGFW.

The performance of genetic and genomic predictions was evaluated using the linear regression (LR) method as described by Legarra and Reverter (2018). The LR method provides a series of statistics derived from the comparison of genetic evaluations using the whole and partial datasets, resulting in easy-to-use methods to evaluate the reliability of the predictions (Legarra & Reverter, 2018). The LR statistics obtained were:

$$\text{Accuracy} = \sqrt{\frac{\text{cov}((G)EBV_w, (G)EBV_p)}{(1 - \bar{F})\sigma_u^2}}, \quad (9)$$

$$\text{Bias} = \text{ave}(\widehat{\mathbf{u}}_{(G)EBV_p}) - \text{ave}(\widehat{\mathbf{u}}_{(G)EBV_w}), \quad (10)$$

$$\text{Dispersion} = \left(\frac{\text{cov}((G)EBV_w, (G)EBV_p)}{\text{var}((G)EBV_p)} \right) - 1, \quad (11)$$

where $\text{cov}((G)EBV_w, (G)EBV_p)$ is the covariance between the GEBV or EBV in whole ($(G)EBV_w$) and partial ($(G)EBV_p$) datasets, \bar{F} is the average inbreeding, $\text{ave}()$ represent the arithmetic average function, $\widehat{\mathbf{u}}_{(G)EBV_w}$ and $\widehat{\mathbf{u}}_{(G)EBV_p}$ are the predicted GEBV or EBV in the whole and partial datasets, respectively, and $\text{var}((G)EBV_p)$ is the variance of the GEBV or EBV. The other components were previously described.

In addition to the LR statistics, the individual theoretical accuracies (TA) were calculated for the focal individuals according to Van Vleck (1993):

$$\text{TA} = \sqrt{1 - \frac{se_i^2}{(1 + f_i)\sigma_u^2}}, \quad (12)$$

where se_i^2 is the square of the GEBV or EBV standard error for the individual i , f_i is the inbreeding coefficient for the individual i , and the other variables were previously described. The TA was calculated for both whole and partial datasets but presented only for the whole dataset to highlight the

overall increase in the TA comparing GEBV and EBV considering all phenotypes available.

2.6 | Evaluated scenarios

The scenarios consisted of combinations of (1) A-BLUP, (2) H-BLUP with $\alpha = 0.95$ and 0.50 to construct \mathbf{G} , and (3) HAP-BLUP using ps-SNP from different LD thresholds (0.15, 0.35, 0.50, 0.65, and 0.80) also with $\alpha = 0.95$ and 0.50 to construct \mathbf{G} . In total, 13 scenarios were evaluated for each of the six traits, resulting in 78 analyses.

3 | RESULTS

3.1 | Genomic prediction accuracies

All the detailed results for the genetic and genomic predictions using the 32K panel including or not the haplotypes are present in the Appendix S2: Tables S1–S6. The prediction accuracies for body weight, wool, and NLB traits ranged from 0.143 (A-BLUP for PWT; Appendix S2: Table S2) to 0.330 (A-BLUP for YGFW; Appendix S2: Table S5). The lowest and highest prediction accuracies were observed for NLB and wool (both YFD and YGFW) traits, respectively. Similar prediction accuracies were observed for the HAP-BLUP across different LD thresholds, regardless of the α value and trait evaluated. We, therefore, only presented the results for the HAP-BLUP considering the LD threshold of 0.50 (HAP-BLUP-LD_0.50) to compare the predictions between pedigree, SNP, and haplotype-based methods for all traits (Figure 1).

Using genomic information provided similar or higher GEBV prediction accuracies compared to EBV, except for BWT and YGFW. An increase of ~41% (~0.06), ~62% (~0.12), ~8% (~0.02), and ~37% (~0.05) in the GEBV prediction accuracies was observed for the PWT, YWT, YFD, and NLB, respectively, when using α equal to 0.95. Using an α of 0.50 generally resulted in half of the increase in the prediction accuracy compared to 0.95. No gains in GEBV accuracy were observed for BWT and YGFW by using genomic information. The increase in the accuracy using the SNP- and haplotype-based models were similar, with differences smaller than 1% for all traits.

3.2 | Bias

The prediction bias ranged between −0.104 (H-BLUP for PWT; Appendix S2: Table S2) and 0.087 (HAP-BLUP using haplotypes from blocks with LD threshold of 0.15 and α of 0.95 for YGFW; Appendix S2: Table S5). Different

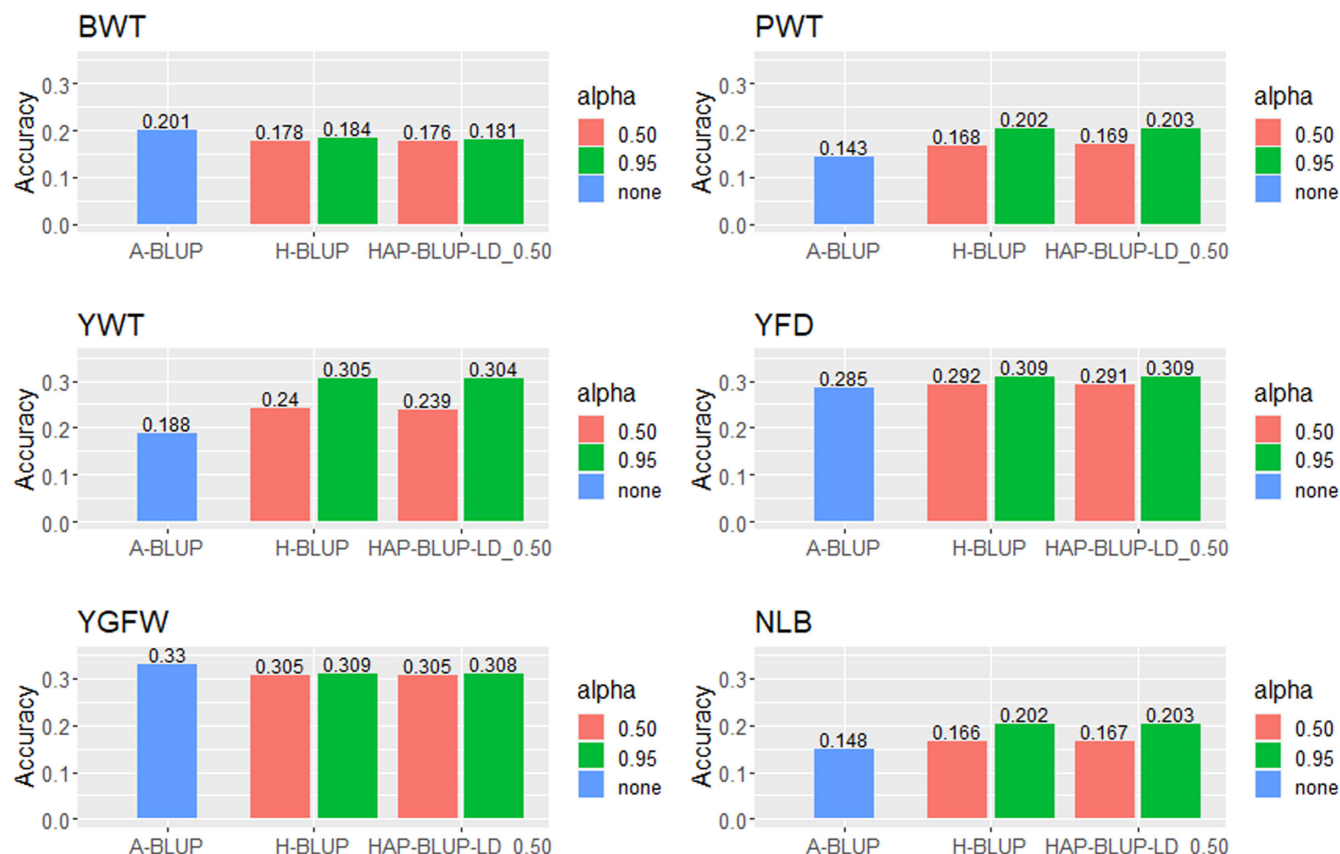


FIGURE 1 Prediction accuracies for birth weight (BWT), postweaning weight (PWT), yearling weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep using pedigree BLUP (A-BLUP), SNP-based single-step GBLUP (H-BLUP), and haplotype-based single-step GBLUP fitting haplotypes from blocks with LD threshold of 0.50 (HAP-BLUP-LD_0.50). Different α values (0.50 or 0.95) were used to create the genomic relationship matrices [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/jbe.12748)]

from what was observed for the GEBV accuracies, the predictions for NLB were in general less biased than the other traits while those for PWT were the most biased. The prediction bias for the haplotype-based methods was similar across LD thresholds (used to create the haploblocks) and, thus, only the HAP-BLUP-LD_0.50 were presented for comparison purposes (Figure 2).

Incorporating genomic information in the analyses resulted in similar or more bias when compared to the pedigree-based prediction. Alpha equal to 0.95 tended to reduce the bias for BWT and YWT, while the opposite was observed for the other traits (i.e., using $\alpha = 0.50$ reduced the prediction bias for the other traits).

3.3 | Dispersion

The GEBV dispersion ranged from -0.572 (A-BLUP for PWT; Appendix S2: Table S2) to 0.035 (A-BLUP for YGFW; Appendix S2: Table S5). The dispersion was closer to zero (expected value for this statistic under no dispersion) for YGFW while it was more distant from (and typically below) zero for BWT and PWT indicating

GEBV were overestimated. GEBV predictions using haplotypes from blocks with different LD thresholds resulted in similar dispersion of the GEBV. Therefore, the HAP-BLUP-LD_0.50 scenario was also used to represent the haplotype-based methods to compare with A-BLUP and H-BLUP (Figure 3).

A dispersion of -0.29 , -0.16 , and -0.35 was observed for PWT, YWT, and NLB, respectively, using H-BLUP and HAP-BLUP-LD_0.50 with α of 0.95. Those values were closer to zero than when using A-BLUP (-0.57 , -0.35 , and -0.39 , respectively), showing reduced dispersion for genomic-based models. Pedigree-based models presented similar or lower dispersion for BWT and wool traits. The dispersion with H-BLUP and HAP-BLUP-LD_0.50 showed similar results regardless of the α values. Alpha equal to 0.95 tended to present better dispersion for PWT, YWT, and NLB compared to 0.50, while the opposite was observed for the other traits.

3.4 | Theoretical accuracy

The mean TA ranged from 0.499 (A-BLUP for PWT; Appendix S2: Table S2) to 0.795 (HAP-BLUP using

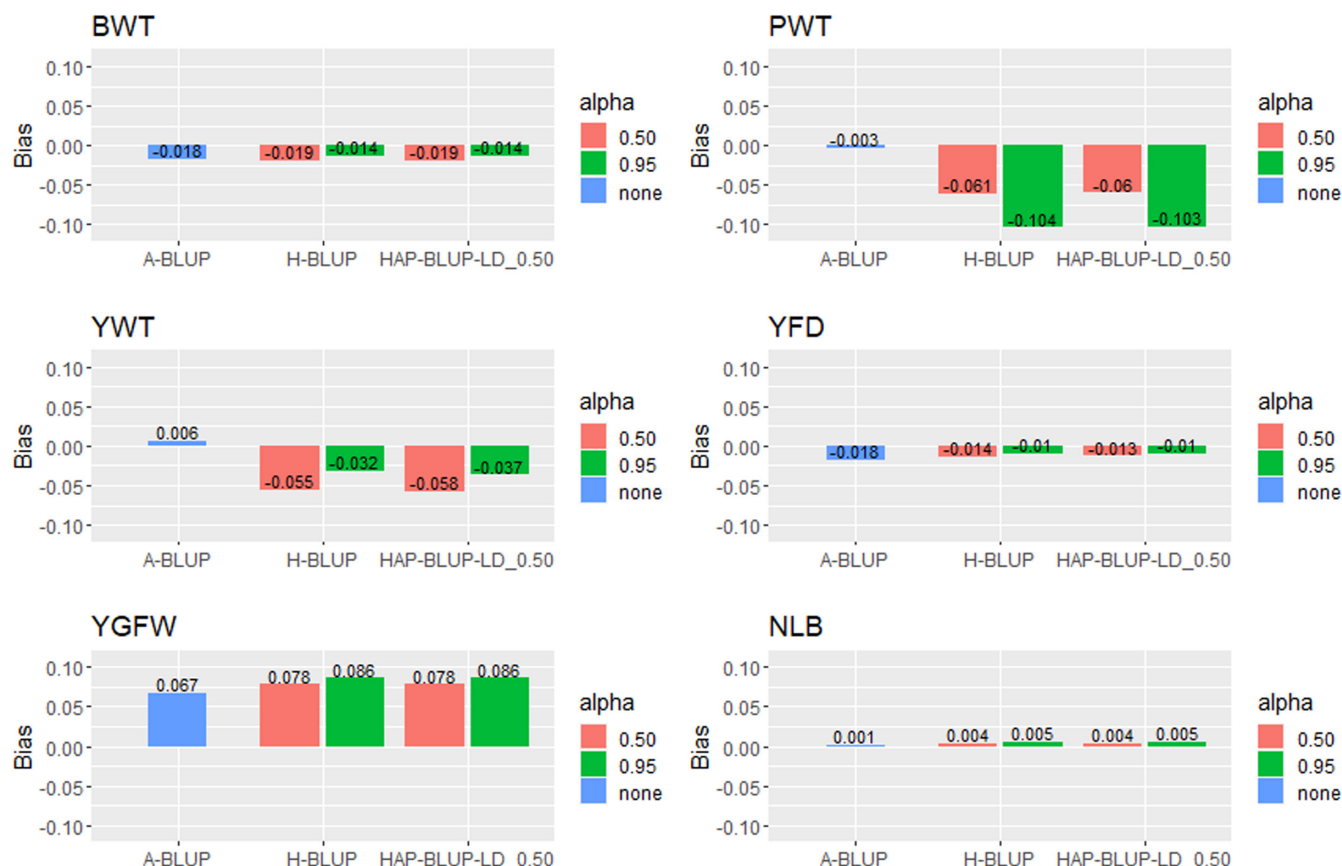


FIGURE 2 Prediction bias for birth weight (BWT), postweaning weight (PWT), yearling weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep using pedigree BLUP (A-BLUP), SNP-based single-step GBLUP (H-BLUP), and haplotype-based single-step GBLUP fitting haplotypes from blocks with LD threshold of 0.50 (HAP-BLUP-LD_0.50). Different α values (0.50 or 0.95) were used to create the genomic relationship matrices [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/jbe.12748)]

haplotypes from blocks with LD threshold of 0.35 and alpha equal to 0.95 for YFD; Appendix S2: Table S4). Considering all traits, the mean TA was 0.631 (0.085) and TA values were higher for YFD and lower for PWT. Results from the haplotype-based methods had similar mean TA regardless of the LD threshold used to construct the haploblocks for all traits. The HAP-BLUP-LD_0.50 was, therefore, again used to represent the HAP-BLUP methods (Figure 4).

The genomic information tended to improve the mean TA for all traits, with increases up to ~7% (~0.04), ~9% (~0.05), ~4% (~0.03), ~3% (~0.02), ~5% (~0.03), and ~6% (~0.04) for BWT, PWT, YWT, YFD, YGFW, and NLB, respectively, using H-BLUP and HAP-BLUP-LD_0.50 with α of 0.95. Negligible difference (less than 1%) was observed in the increase of the mean TA between H-BLUP and HAP-BLUP-LD_0.50 with α of 0.95. Using α equal to 0.50 resulted in the smallest increase in the mean TA (less than 2%) with both SNP- and haplotype-based methods for all traits. At the individual level, the TA using H-BLUP with α of 0.95 were higher compared to A-BLUP for the younger individuals and those

with no phenotypic information (sires and dams with genotyped progeny) in the partial datasets for all traits (Figure 5).

Increase in the TA for ungenotyped close relatives (progeny or parents) of the genotyped individuals when using genomic information was also observed (Appendix S3: Figure S1), although the increases were smaller than those for the genotyped individuals. In the case of ungenotyped unrelated (animals with no pedigree links) individuals to the genotyped cohort, most of the TA with or without genomics were similar (Appendix S3: Figure S2); only few of these unrelated individuals benefited from the use of genomic information.

4 | DISCUSSION

Genomic selection is the state-of-the-art in modern sheep breeding programs. Here we present the first results of genomic predictions for body weight, wool, and reproductive traits in North American Rambouillet

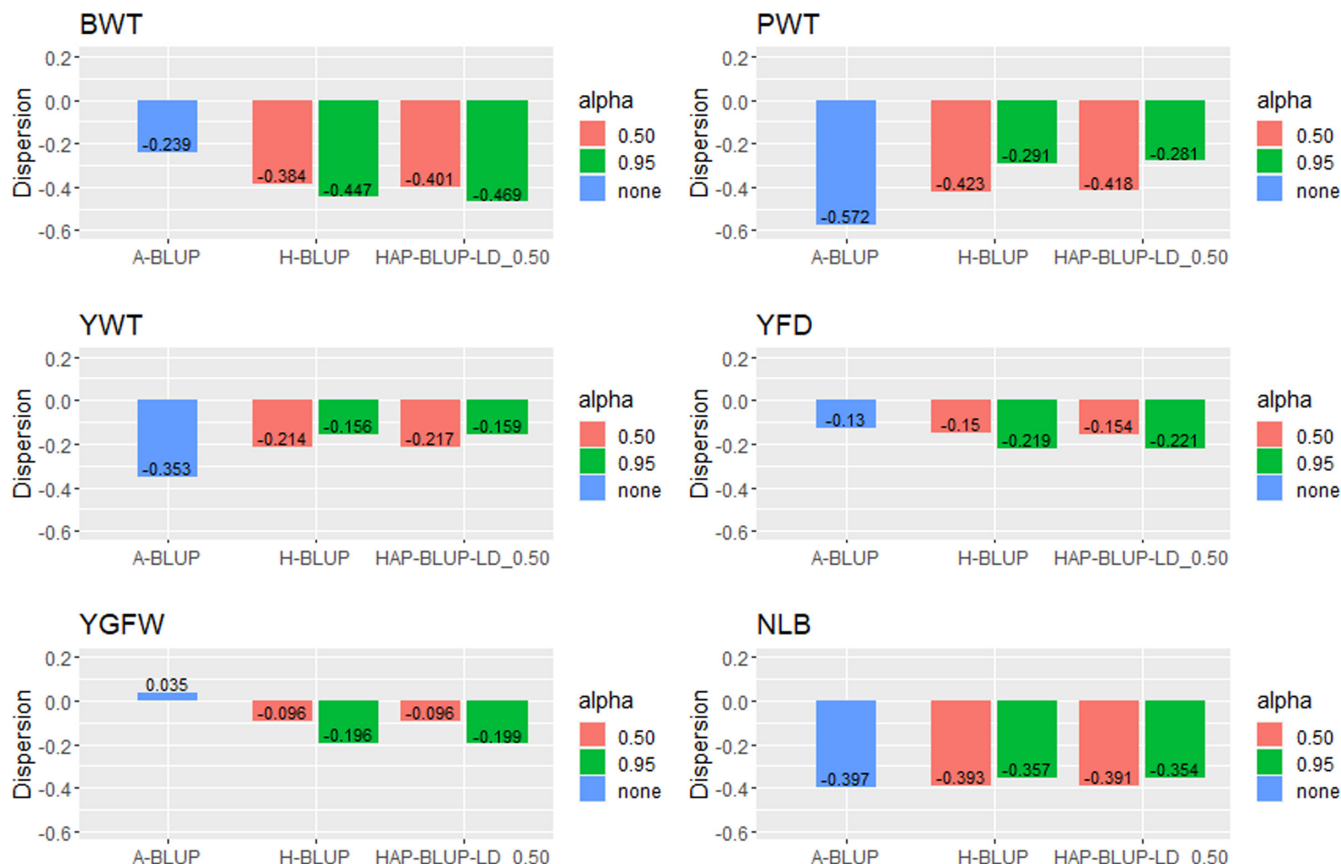


FIGURE 3 Dispersion of the genomic-enhanced breeding values (GEBV) for birth weight (BWT), postweaning weight (PWT), yearling weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep using pedigree BLUP (A-BLUP), SNP-based single-step GBLUP (H-BLUP), and haplotype-based single-step GBLUP using haplotypes from blocks with LD threshold of 0.50 (HAP-BLUP-LD_0.50). Different α values (0.50 or 0.95) were used to create the genomic relationship matrices [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

sheep. We performed single-step genomic predictions fitting SNP or haplotypes to create the genomic relationship matrices used to compute GEBV. Despite the small dataset, promising results were obtained, which can be used as a starting point for the implementation of genomic selection in Rambouillet sheep as well as in other sheep populations.

4.1 | Genetic and genomic prediction results

The accuracy of genomic predictions relies mainly on the trait heritability, LD between SNP and QTL (Meuwissen et al., 2001), population structure, and genetic diversity of the population (Daetwyler et al., 2012). For Rambouillet sheep, the prediction accuracies for the pedigree- and genomic-based models followed the expected pattern in being higher for more heritable traits (Table 1; Figure 1). Correspondingly, the smallest (~0.02 in absolute value, ~8%) differences between

the pedigree- and genomic-based prediction accuracies were observed for YFD, which was the trait with the highest heritability (0.57), in comparison to NLB (~0.05 in absolute value, ~37%)—the trait with lowest heritability (0.09).

Despite the expectation of the theoretical accuracies of genomic predictions to be higher for traits with higher heritability (Meuwissen et al., 2001), benefits of using GEBV are expected to be higher for traits with low heritability, sex-limited, hard-to-measure, and recorded late in life, especially in sheep (Brito, Clarke, et al., 2017; Brown et al., 2018). For more highly heritable and easily measured traits, given that the assumptions for the MME (Henderson, 1984) are met (e.g., polygenic architecture, deep and accurate pedigree data, large number of phenotypic records, no preselection), the EBV are expected to be BLUP and predict the unknown true breeding values well. In other words, the prediction accuracies using A-BLUP for highly heritable traits where individuals and/or their progeny have phenotypes are already expected to be high.

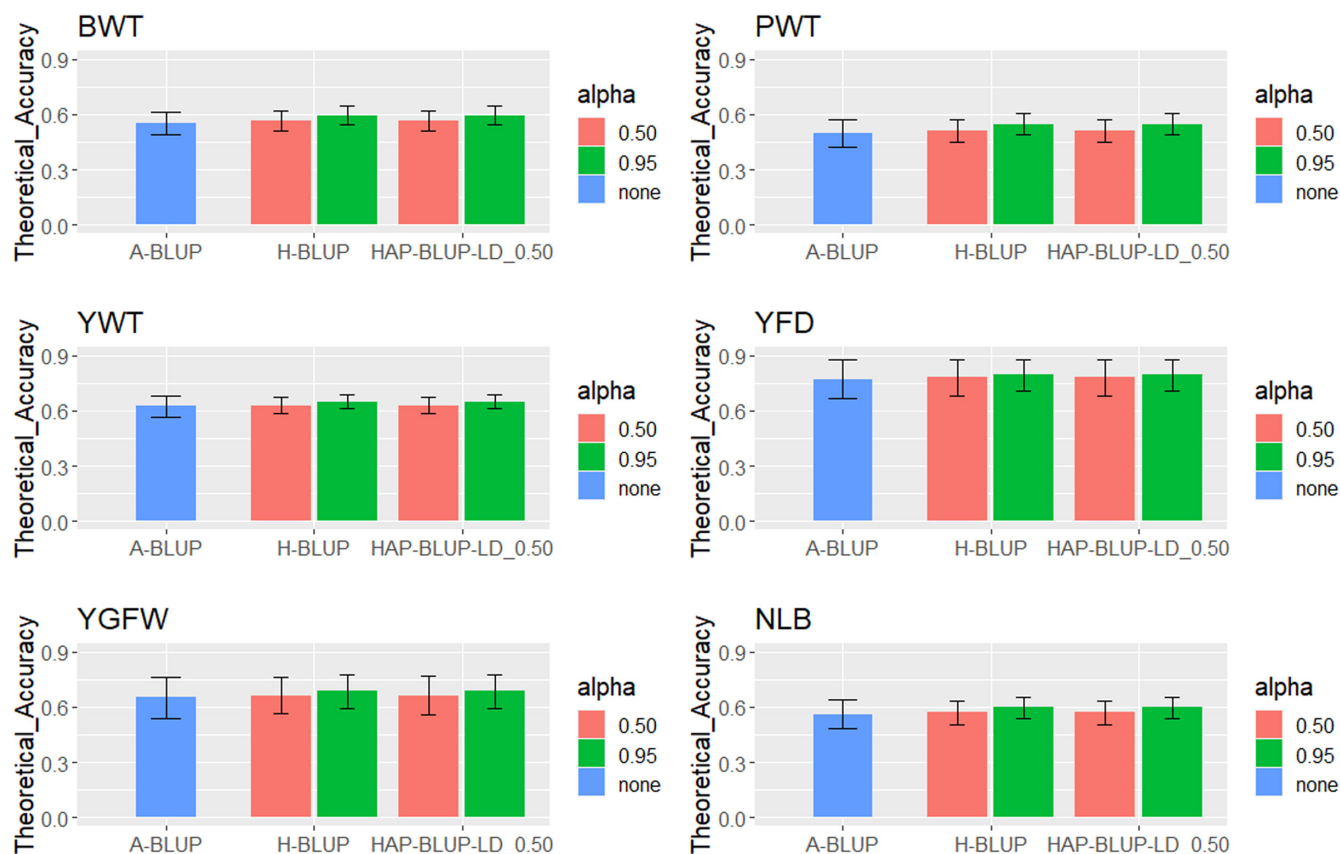


FIGURE 4 Mean theoretical accuracies for birth weight (BWT), postweaning weight (PWT), yearling weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB) in Rambouillet sheep using pedigree BLUP (A-BLUP), SNP-based single-step GBLUP (H-BLUP), and haplotype-based single-step GBLUP using haplotypes from blocks with linkage disequilibrium (LD) threshold of 0.50 (HAP-BLUP-LD_0.50). Different α values (0.50 or 0.95) were used to create the genomic relationship matrices [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/jbr.12748)]

As the genotyped cohort used in this study is the first official attempt to create a training population for genomic evaluations of Rambouillet sheep, the genotyped individuals included key ancestors and other selected animals with phenotypes or representative progeny with phenotypes for the traits evaluated. In this case, the EBV for those individuals were well estimated, especially for YFD and YGFW (higher heritabilities; Table 2), with small and negligible increase in the prediction accuracy for the genomic- compared to pedigree-based models. For such highly heritable traits, using GEBV would still be more important to select breeding candidates at a younger age, that is, measured only at yearling age. The substantial increase in the prediction accuracy of PWT, YWT, and NLB (higher than ~37%) shows that greater genetic gains can be achieved for these traits by including genomic information, as the accuracy is one of the main components of the selection response (Falconer & Mackay, 1996).

The genomic prediction accuracies observed in our study were within the range for most of the economic traits in sheep, which is between 0.20 and 0.50 according to Brown et al. (2018), especially when using α of 0.95 to

construct the **G** matrix. Oliveira et al. (2021) observed prediction accuracies for BWT ranging from 0.06 to 0.13 using H-BLUP for Norwegian White and New Zealand Composite sheep populations. Unlike what was observed in the current study, Moghaddar et al. (2019) showed accuracies for genomic predictions ranging between 0.40 to 0.60 for PWT, 0.30 to 0.40 for yearling clean fleece weight, and 0.30 to 0.50 for YFD using BayesR and GBLUP for purebred Merino and crosses between Merino and Border Leicester. Genomic prediction accuracies of 0.24 and 0.28 were observed for YGFW using GBLUP and BayesR, respectively, and 0.31 and 0.35 for YFD for the same methods, respectively, in Merino and crossed Merino (Bolormaa, Brown, et al., 2017). For NLB, Bolormaa, Swan, et al. (2017) reported genomic prediction accuracies ranging from 0.15 to 0.56 considering different validation strategies and prediction based on GBLUP and BayesR under cross-validation approaches for Merino sheep. Those differences in the genomic prediction accuracies for the same traits are also related to the statistical model and validation method used in the evaluations, beyond the other factors previously mentioned (e.g., heritability, population

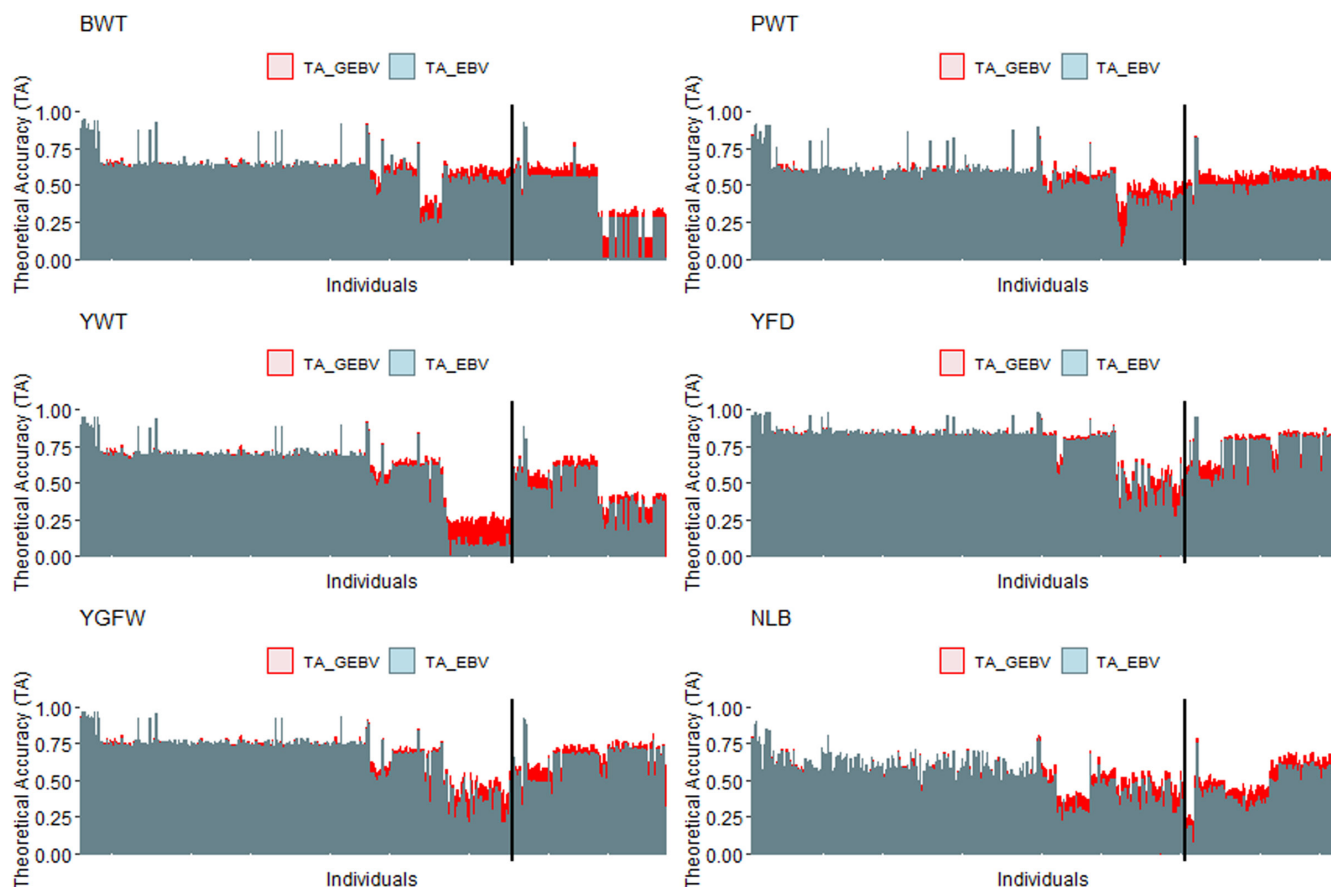


FIGURE 5 Theoretical accuracies for the genomic estimated breeding values using SNP (TA_GEBV) and estimated breeding values (TA_EBV) per genotyped individuals for birth weight (BWT), postweaning weight (PWT), yearling weight (YWT), yearling fibre diameter (YFD), yearling greasy fleece weight (YGFW), and number of lambs born (NLB). The TA_GEBV and TA_EBV were obtained using SNP in the single-step GBLUP (H-BLUP) with alpha equal to 0.95 and pedigree-based BLUP (A-BLUP), respectively. The individuals were sorted by birth date, so that the youngest individuals are in the right side of each plot. The black vertical line divides the training individuals in the partial datasets, with the focal animals on the right side of the line [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/jbe.12748)]

structure, and genetic diversity). Also, GEBV accuracies were not calculated in the same way across all the studies, but the LR method used in this study is currently considered as the gold-standard approach.

The regression coefficient of the adjusted (or corrected) phenotypes or EBV on the GEBV is usually used to measure the “bias” of GEBV (Brown et al., 2018; Gao et al., 2012; Moghaddar et al., 2019; Oliveira et al., 2021). This measure was assessed as dispersion in our study, as it represents how the GEBV were inflated or deflated (over- or under-estimated, respectively). Prediction bias, as a property of the method and the population under evaluation, is the expectation of the difference between average true and predicted breeding value; bias is zero under ideal conditions and can be approximated by the difference between the average (G)EBV in the whole and partial data sets (Legarra & Reverter, 2018). The observation that GEBV and EBV for most of the scenarios across traits were over-estimated in the Rambouillet sheep was consistent with other studies (Brown et al., 2018; Moghaddar et al., 2019;

Oliveira et al., 2021). Our conclusions regarding the benefit of including genomic information in prediction based on dispersion followed the same pattern observed for the GEBV prediction accuracies; this was likely because they are affected by similar factors.

Reports of prediction bias in sheep, as described by Legarra and Reverter (2018), are scarce and were found only in dairy sheep (Macedo et al., 2020, 2022). The same is true for TA, although this is an important metric when reporting breeding values back to producers. Brito, Clarke, et al. (2017) observed TA values ranging from 0.25 to 0.49 across a range of growth, carcass, and meat quality traits, which are smaller than the values observed in the current study. The substantial increase in the TA (Figure 4) especially for the young individuals using genomic information (Figure 5) is promising because these are the key individuals that need to be ranked for selection purposes.

Importantly, ungenotyped close relatives of the genotyped animals also had greater TA when incorporating genomic information. This suggests that genotyping

strategies considering family information could also be used to improve the TA for ungenotyped individuals through genetic links. Similar to this study, Massender et al. (2022) also found that including genomic information provided higher TA for genotyped individuals compared to pedigree-based predictions in dairy goats, with increases for ungenotyped animals with genotyped relatives. Regular and accurate pedigree collection is therefore encouraged to establish genetic links and evaluate which ungenotyped individuals could benefit more from having genomic information. An improved pedigree collection is also likely to increase the pedigree depth and quality, being important to control the genetic diversity of the population, which affects the long-term responses of the breeding program (Stachowicz et al., 2018). The PCI is an important measure of pedigree quality and the average observed in this study (0.57) was higher than those for Canadian Suffolk (0.48) and for the major Canadian sheep breeds combined (0.50) reported by Stachowicz et al. (2018); nevertheless, the higher the PCI the better.

Selective genotyping can result in maximum genetic response (Boligon et al., 2012), which could explain the improvements in the prediction results for most of the traits analysed using genomic information. Such was the case even when the predictions were based on a small number of individuals (242 to 632 for NLB and BWT, respectively) genotyped using a moderate SNP density panel (~32 K SNP). Most of the prediction accuracy using genomic information is due to population structure, as described by Daetwyler et al. (2012). Those authors showed that up to 86% of the prediction accuracy can be achieved by using only one chromosome in a multibreed sheep population. Although one chromosome was enough to capture the population structure, it was unlikely to contain all the QTL affecting a trait (Daetwyler et al., 2012). Nevertheless, the recommendation is to increase the SNP panel density, through genotyping and imputation, for genomic predictions so that both population structure and LD between marker and QTL are fully explored (Daetwyler et al., 2012). Evaluating weighted single-step genomic predictions and genome-wide associations has also been encouraged (Wang et al., 2012) as there may be important genomic regions that explain more of the total additive genetic variance for the traits of interest.

Selective genotyping can increase bias in the variance component estimation and therefore is not recommended for the breeding programs that only use phenotypes and pedigree relationships to drive selection decisions (Wang et al., 2020). The potential bias from selective genotyping was a reason for using the variance components provided by the NSIP, derived using solely phenotypic and pedigree information. Using random selection to choose the samples to be genotyped, as well as increasing the training

population size, is recommended to avoid bias in both variance component estimation and GEBV, as more biased predictions were observed using genomic information for most of the traits (Figure 2).

4.2 | Using different alpha values to construct the genomic relationship matrices

In general, appropriate α and β parameters have more impact in GEBV bias reduction (Gao et al., 2012). Despite greater GEBV accuracies for a higher (0.95) α value, 0.50 is the choice to create **G** in single-step evaluations for a range of carcass traits in terminal sire sheep breeds in Australia (McMillan & Swan, 2017). According to these authors, an α equal to 0.50 was chosen because (1) when increasing α , accuracies increased until reaching an asymptote at around 0.50, which was not the case in the current study; (2) the GEBV using α between 0.50 and 0.95 were highly correlated; and (3) less variation was observed in the GEBV of genotyped individuals without phenotypes with α equal to 0.50; and (4) with higher α values, GEBV bias (over-prediction) increased.

In this study, using α of 0.50 showed only half of the increase in the accuracies compared to 0.95. No clear advantage in GEBV bias, dispersion, or TA was observed with one α value compared to the other. Therefore, we recommend α of 0.95 for single-step genomic evaluations in U.S. Rambouillet sheep. However, it is important to highlight that our study included a smaller number of genotyped individuals compared to McMillan and Swan (2017) and we used the LR method (Legarra & Reverter, 2018) to derive the GEBV accuracies, bias, and dispersion; they instead used cross-validation to test their predictions with random assignment of individuals to groups. Furthermore, it is recommended to evaluate the impact of different α and β in the creation of genomic relationships in multiple trait models in the future, as these parameters can also change for different traits (Gao et al., 2012; McMillan & Swan, 2017) and pedigree quality.

4.3 | Haplotype-based single-step genomic predictions

The HAP-BLUP-LD_0.50 scenario was chosen to represent the haplotype-based methods because the LD of 0.50 was the level most likely to estimate the recombination hotspots properly, which are the specific points in the genome with higher probability of recombination (Kim et al., 2018). The use of haplotype-based methods did not improve the prediction results for any of the traits

analysed (accuracies, bias, dispersion, and TA) compared to fitting SNP in a real (as opposed to simulated data) sheep dataset. This may have been the consequence of not enough changes in the genomic relationships to result in differences in the GEBV to be more accurate, as the correlations between the **G** matrices using only SNP or SNP plus haplotypes were high (~0.99). Haplotype-based genomic predictions could outperform SNP-based models in sheep datasets because the former can capture epistasis and these populations could have more complex interactions within haplotype blocks due to higher effective population size. Liang et al. (2020) showed that epistasis was the main reason for higher GEBV accuracies when using haplotypes instead of SNP in seven traits in humans, which is a highly genetically diverse population (Park, 2011). However, in this study, the small number of genotyped individuals ($n = 722$) as well as the density of the SNP panel used (~32 K SNP) could have affected both SNP and haplotype genomic predictions.

The algorithm to create the LD-based haploblocks and the method to code the haplotypes during the creation of the relationship matrix could also have affected the prediction results. There are several algorithms to create LD-haploblocks, such as MATILDE (Pattaro et al., 2008), confidence interval (Gabriel et al., 2002), four gamete test (Wang et al., 2002), solid spine (Barrett et al., 2005), MIG++ (Taliun et al., 2014), S-MIG++ (Taliun et al., 2016), and Big-LD (Kim et al., 2018). We have used the Big-LD algorithm to construct the haploblocks because the LD blocks produced by this method agree better with the true recombination hotspots (determined experimentally in the major histocompatibility complex region from semen of north-European British donors) and are more computationally efficient than the previously mentioned algorithms (Kim et al., 2018). However, as the haplotype diversity index and true discovery ratio of the recombination hotspots can be lower using Big-LD (Kim et al., 2018), evaluating different algorithms to create the LD-haplotype blocks is also recommended. New methods based on clustering algorithms (Won et al., 2020) and machine learning methods (Lim et al., 2022) have been recently proposed to create and select the best haplotypes to be used, respectively. These approaches should be evaluated in future studies with larger genomic datasets and populations with different genetic backgrounds.

The haplotypes can be multiallelic markers (Calus et al., 2008; Gabriel et al., 2002). However, we used the unique multiallelic haplotype alleles coded as ps-SNP to perform genomic predictions under the ssGBLUP framework. The ps-SNP derived from the LD-haploblocks were then merged with the NCSNP to create the **G** matrix, similar to Araujo et al. (2021). This strategy enables fitting haplotypes to perform genomic predictions using software

developed for fitting individual SNP (Teissier et al., 2020), including or excluding non-genotyped individuals (ssGBLUP and GBLUP, respectively). Teissier et al. (2020) considered both NCSNP and unique multiallelic haplotype alleles as ps-SNP and observed up to 22% increase in GEBV prediction accuracy when using different LD- or fixed-SNP-length-based haploblocks for milk production traits in dairy goats using ssGBLUP. Milk production traits are known to be affected by a major gene (*DGATI*). In general, GEBV prediction results for haplotype-based methods are scarce in small ruminants and additional studies are needed.

The GVCHAP is a computing pipeline that allows multiallelic haplotypes to be used directly to create a genomic additive (and dominance) relationship matrix for both genomic prediction and variance component estimation using haplotypes or SNP (Prakapenka et al., 2020). GVCHAP is based in the multiallelic haplotype model proposed by Da (2015), which uses the quantitative genetic theory to derive a general multiallelic partition of genotypic values with factorization to define the genomic relationships. However, the GVCHAP is based on GREML and GBLUP and, thus, only considers genotyped individuals with phenotypes. Considering the different algorithms and methods to create haploblocks, code the haplotype alleles, and create the genomic relationship matrix including haplotypes, there are still further alternatives to evaluate the feasibility of including haplotypes in genomic predictions. Future studies in sheep could also consider the possibility of creating haplotypes based on functional information (e.g., gene regions) to perform haplotype predictions (Da, 2015; Prakapenka et al., 2020).

Despite the hypothesis that haplotypes could outperform SNP and provide high accuracies and lower bias in genomic predictions, recent results have shown that this does not usually happen. As summarized by Araujo et al. (2021), the benefits of using haplotype-based methods for genomic prediction are equivocal most of the time. Significant improvements usually occur mainly in the evaluation of traits with major genes, as shown by Teissier et al. (2020). Nevertheless, as stated before, there are haplotype blocking and selection methods that should be further investigated. There are methods that might be able to better capture the haplotype variation and improve GEBV accuracy.

Marker density can also affect the accuracy of SNP phasing (Weng et al., 2014) and the precision in which the recombination hotspots are determined (Weng et al., 2019), which are the first steps for the haplotype prediction and the basis of the LD-based haploblocks, respectively. In addition, epistasis, which is the component that might contribute the most to improvements in accuracy with haplotype predictions (Liang et al., 2020), is a complex effect and requires a

substantial number of individuals and markers/bins to be properly estimated (Zhang et al., 2016).

Despite the fact that we genotyped individuals for SNP panel with a higher density (50K and 600K) than the one used in this research (35K), genotype imputation prior to haplotype blocking and prediction can also be hampered by the number of reference individuals (Ventura et al., 2016). In early stages of this research, we imputed the 50K panel from the overlapping 35K SNP panel for those animals with HD genotypes. The results of genomic predictions using the imputed 50K for both SNP and haplotype-based methods (Appendix S4: Tables S1–S6) were similar to that using only the overlapping ~35 K SNPs, except for the YGFW (Appendix S4: Table S5). This likely happened because increasing the SNP density of only ~60 individuals by ~6 K SNP (imputed variants that passed a QC similar to that described in the section 2.2) to create the **G** matrix to be combined with **A** to create the **H**⁻¹ was not large enough to result in significant improvements in the genomic predictions. The imputation accuracy for SNP and individuals was ~80% and ~77%, respectively, which were below the 90% threshold suggested by Bolormaa et al. (2015).

Even though the accuracy of predictions for YGFW using haplotypes based on imputed 50K panels was higher (up to ~0.09 compared to predictions using the 35K panel), generally predictions using the imputed 50K were more biased and the GEBV were more dispersed for all traits. Legarra and Reverter (2018) emphasized that accuracies should only be estimated and compared from empirically unbiased models. The results of the LR statistics using the imputed set of markers could have been affected by bias because of the imputation process due to limitations in the size of the reference population. Our results based on the 35K SNP panel should therefore be used considering the Rambouillet sheep currently genotyped. Larger reference populations and denser SNP panels are recommended to evaluate genomic predictions using haplotypes in future sheep studies.

5 | CONCLUSIONS

Using genomic information (SNP or haplotypes as pseudo-SNP) provided similar or higher GEBV prediction and theoretical accuracies, and reduced the dispersion of the GEBV, for body weight, wool, and reproductive traits in Rambouillet sheep. However, there were no clear improvements in prediction bias compared to pedigree-based predictions. Alpha value equal to 0.95 is recommended to weight the genomic relationships when modelling the covariances between individuals. The use of haplotypes showed no advantage compared to SNP at the current reference population size and SNP panel density used, regardless of the LD threshold used to create

the haploblocks. SNP-based genomic predictions are therefore recommended since they are easier to implement than those based on haplotypes. Efforts to increase the number of genotyped individuals are paramount to take full advantage of genomic information to accelerate genetic progress in the U.S. Rambouillet sheep breeding program.

AUTHOR CONTRIBUTION

ACA, PLSC, RML, and LFB: conception of the work; RML and LFB: data acquisition; ACA: SNP and haplotype-based single-step genomic prediction analyses; ACA, RML, and LFB: results interpretation; ACA: drafting the manuscript; ACA, PLSC, HRO, RML, and LFB: critical revision of the manuscript; and, ACA, PLSC, HRO, RML, and LFB: final approval of the version to be published. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

We acknowledge the NSIP for providing the datasets used in this research, and the nine NSIP Rambouillet flocks that provided the DNA samples genotyped. We also thank Dr. Jessica Petersen (University of Nebraska-Lincoln) for updating the marker positions to the most recent ovine reference genome (ARS-UI Ramb v2.0) and merging the SNP panels used, Dr. Mathew Spangler (University of Nebraska-Lincoln) for providing useful comments on the final version of the manuscript, and other members of the Brito's lab for providing scientific support to develop this research. Lastly, we thank Purdue University and the State University of Southwest Bahia for providing academic and financial support to the authors. We also acknowledge the National Development Council Scientific Technological (CNPq, Brazil) for the fellowship.

FUNDING INFORMATION

This research was funded by the National Sheep Industry Improvement Center (NSIIC-USDA), the American Rambouillet Sheep Breeders Association, and the American Sheep Industry Association Let us Grow program. The State University of Southwest Bahia and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil), Financial Code 001, provided the scholarship of the first author.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The phenotypic, pedigree and genomic data used in this study are the property of the industry partner that

contributed to the study and therefore are not readily available due to its commercial sensitivity. Requests to access the datasets should be directed to the National Sheep Improvement Program (NSIP). The computing pipelines used in this research are available by request to the corresponding authors.

ORCID

Andre C. Araujo  <https://orcid.org/0000-0002-8057-472X>

Luiz F. Brito  <https://orcid.org/0000-0002-5819-0922>

REFERENCES

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final. *Journal of Dairy Science*, 93(2), 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Araujo, A. C., Carneiro, P. L. S., Alvarenga, A. B., Oliveira, H. R., Miller, S. P., Retallick, K., & Brito, L. F. (2022). Haplotype-based single-step GWAS for yearling temperament in American Angus cattle. *Genes*, 13(17), 1–27. <https://doi.org/10.3390/genes13010017>
- Araujo, A. C., Carneiro, P. L. S., Oliveira, H. R., Schenkel, F. S., Veroneze, R., Lourenco, D. A. L., & Brito, L. F. (2021). A comprehensive comparison of haplotype-based single-step genomic predictions in livestock populations with different genetic diversity levels: A simulation study. *Frontiers in Genetics*, 12(729867), 1–17. <https://doi.org/10.3389/fgene.2021.729867>
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–265. <https://doi.org/10.1093/bioinformatics/bth457>
- Boligon, A. A., Long, N., Albuquerque, L. G., Weigel, K. A., Gianola, D., & Rosa, G. J. M. (2012). Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection. *Journal of Animal Science*, 90(13), 4716–4722. <https://doi.org/10.2527/jas.2012-4857>
- Bolormaa, S., Brown, D. J., Swan, A. A., van der Werf, J. H. J., Hayes, B. J., & Daetwyler, H. D. (2017). Genomic prediction of reproduction traits for merino sheep. *Animal Genetics*, 48(3), 338–348. <https://doi.org/10.1111/age.12541>
- Bolormaa, S., Gore, K., van der Werf, J. H. J., Hayes, B. J., & Daetwyler, H. D. (2015). Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Animal Genetics*, 46(5), 544–556. <https://doi.org/10.1111/age.12340>
- Bolormaa, S., Swan, A. A., Brown, D. J., Hatcher, S., Moghaddar, N., van der Werf, J. H. J., Goddard, M. E., & Daetwyler, H. D. (2017). Multiple-trait QTL mapping and genomic prediction for wool traits in sheep. *Genetics Selection Evolution*, 49(62), 62. <https://doi.org/10.1186/s12711-017-0337-y>
- Bovo, S., Ballan, M., Schiavo, G., Ribani, A., Tinarelli, S., Utzeri, V. J., Dall'Olio, S., Gallo, M., & Fontanesi, L. (2021). Single-marker and haplotype-based genome-wide association studies for the number of teats in two heavy pig breeds. *Animal Genetics*, 52(4), 440–450. <https://doi.org/10.1111/age.13095>
- Brito, L. F., Clarke, S. M., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., Dodds, K. G., Sargolzaei, M., & Schenkel, F. S. (2017). Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genetics*, 18(7), 7. <https://doi.org/10.1186/s12863-017-0476-8>
- Brito, L. F., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., Dodds, K. G., Schenkel, F. S., & Clarke, S. M. (2017). Genetic diversity of a New Zealand multi-breed sheep population and composite Breed's history revealed by a high-density SNP Chip. *BMC Genetics*, 18(25), 25. <https://doi.org/10.1186/s12863-017-0492-8>
- Brito, L. F., Oliveira, H. R., McConn, B. R., Schinckel, A. P., Arrazola, A., Marchant-Forde, J. N., & Johnson, J. S. (2020). Large-scale phenotyping of livestock welfare in commercial production systems: A new frontier in animal breeding. *Frontiers in Genetics*, 11(793), 1–32. <https://doi.org/10.3389/fgene.2020.00793>
- Brown, D. J., Swan, A. A., Boerner, V., Li, L., Gurman, P. M., McMillan, A. J., van der Werf, J. H. J., Chandler, H. R., Tier, B., & Banks, R. G. (2018). Single-step genetic evaluations in the Australian sheep industry. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, 11, 460.
- Calus, M. P. L., Meuwissen, T. H. E., de Roos, A. P. W., & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define Haplotypes fine haplotypes. *Genetics*, 178(1), 553–561. <https://doi.org/10.1534/genetics.107.080838>
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42(1), 1–8. <https://doi.org/10.1186/1297-9686-42-2>
- Cuyabano, B. C., Su, G., & Lund, M. S. (2014). Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. *BMC Genomics*, 15(1171), 1–11. <https://doi.org/10.1186/1471-2164-15-1171>
- Cuyabano, B. C., Su, G., & Lund, M. S. (2015). Selection of haplotype variables from a high-density marker map for genomic prediction. *Genetics Selection Evolution*, 47(61), 1–11. <https://doi.org/10.1186/s12711-015-0143-3>
- Da, Y. (2015). Multi-allelic haplotype model based on genetic partition for genomic prediction and variance component estimation using SNP markers. *BMC Genetics*, 16(144), 144. <https://doi.org/10.1186/s12863-015-0301-1>
- Daetwyler, H. D., Kemper, K. E., van der Werf, J. H. J., & Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal Animal Science*, 90(10), 3375–3384. <https://doi.org/10.2527/jas2011-4557>
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (p. 4). Longman.
- Feitosa, F. L. B., Pereira, A. S. C., Amorim, S. T., Peripolli, E., Silva, R. M. D. O., Braz, C. U., Ferrinho, A. M., Schenkel, F. S., Brito, L. F., Espigolan, R., Albuquerque, L. G., & Baldi, F. (2020). Comparison between haplotype-based and individual SNP-based genomic predictions for beef fatty acid profile in Nelore cattle. *Journal of Animal Breeding and Genetics*, 137(5), 468–476. <https://doi.org/10.1111/jbg.12463>
- Feitosa, F. L. B., Pereira, A. S. C., Mueller, L. F., de Souza Fonseca, P. A., Braz, C. U., Amorim, S., Espigolan, R., Lemos, M. A., de Albuquerque, L. G., Schenkel, F. S., Brito, L. F., Stafuzza, N. B., & Baldi, F. (2021). Genome-wide association study for beef fatty acid profile using haplotypes in

- Nellore cattle. *Livestock Science*, 245, 104396. <https://doi.org/10.1016/j.livsci.2021.104396>
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordeiro, S. N., Rotini, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. L., Daly, M. J., & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225–2229. <https://doi.org/10.1126/science.1069424>
- Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., & Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution*, 44(8), 1–8. <http://www.gsejournal.org/content/44/1/8>
- Guarini, A. R., Lourenco, D. A. L., Brito, L. F., Sargolzaei, M., Baes, C. F., Miglior, F., Misztal, I., & Schenkel, F. S. (2018). Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. *Journal of Dairy Science*, 101(9), 8076–8086. <https://doi.org/10.3168/jds.2017-14193>
- Henderson, C. R. (1984). *Application of linear models in animal breeding*. University of Guelph.
- Hess, M., Druet, T., Hess, A., & Garrick, D. (2017). Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. *Genetics Selection Evolution*, 49(54), 54. <https://doi.org/10.1186/s12711-017-0329-y>
- Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(226), 226–231. <https://doi.org/10.1007/BF01245622>
- Jiang, Y., Schmidt, R. H., & Reif, J. C. (2018). Haplotype-based genome-wide prediction models exploit local Epistatic interactions among markers. *Genes Genomes Genetics*, 8(5), 1687–1699. <https://doi.org/10.1534/g3.117.300548>
- Kennedy, B. W., & Trus, D. (1993). Considerations on genetic connectedness between management units under an animal model. *Journal of Animal Science*, 71(9), 2341–2352. <https://doi.org/10.2527/1993.7192341x>
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto-Neto, L. R., Cristobal, M. S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., & Dalrymple, & other members of the International Sheep Genomics Consortium. (2012). Genome-wide analysis of the World's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, 10, e1001258. <https://doi.org/10.1371/journal.pbio.1001258>
- Kim, S. A., Brossard, M., Roshandel, D., Paterson, A. D., Bull, S. B., & Yoo, Y. J. (2019). Gpart: Human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. *Bioinformatics*, 35(4419), 4419–4421. <https://doi.org/10.1093/bioinformatics/btz308>
- Kim, S. A., Cho, C. S., Kim, S. R., Bull, S. B., & Yoo, Y. J. (2018). A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. *Bioinformatics*, 34(388), 388–397. <https://doi.org/10.1093/bioinformatics/btx609>
- Kuehn, L. A., Lewis, R. M., & Notter, D. R. (2007). Managing the risk of comparing estimated breeding values across flocks or herds through connectedness: A review and application. *Genetics Selection Evolution*, 39(225), 225–247. <https://doi.org/10.1186/1297-9686-39-3-225>
- Kuehn, L. A., Notter, D. R., Nieuwhof, G. J., & Lewis, R. M. (2008). Changes in connectedness over time in alternative sheep sire referencing schemes. *Journal of Animal Science*, 86(3), 536–544. <https://doi.org/10.2527/jas.2007-0256>
- Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9), 4656–4663. <https://doi.org/10.3168/jds.2009-2061>
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single step, a general approach for genomic selection. *Livestock Science*, 166, 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>
- Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, 50(53), 53. <https://doi.org/10.1186/s12711-018-0426-6>
- Lewis, R. M., Crump, R. E., Kuehn, L. A., Simm, G., & Thompson, R. (2005). Assessing connectedness in across-flock genetic evaluations. *Journal of Animal Science*, 83(suppl. 1), 101. <https://doi.org/10.1017/S1752756200002763>
- Liang, Z., Tan, C., Prakapenka, D., Ma, L., & Da, Y. (2020). Haplotype analysis of genomic prediction using structural and functional genomic information for seven human phenotypes. *Frontiers in Genetics*, 11(1), 1–20. <https://doi.org/10.3389/fgene.2020.588907>
- Lim, A. J. W., Lim, L. J., Ooi, B. N. S., Koh, E. T., Tan, J. W. L., Chong, S. S., Khor, C. C., Tucker-Kellogg, L., Leong, K. P., & Lee, C. G. (2022). Functional coding haplotypes and machine-learning feature elimination identifies predictors of methotrexate response in rheumatoid arthritis patients. *eBioMedicine*, 75(103800), 1–18. <https://doi.org/10.1016/j.ebiom.2021.103800>
- Macedo, F. L., Astruc, J. M., Meuwissen, T. H. E., & Legarra, A. (2022). Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. *Journal of Dairy Science*, 105(3), 2439–2452. <https://doi.org/10.3168/jds.2021-20860>
- Macedo, F. L., Christensen, O. L., Astruc, J.-M., Aguilar, I., Masuda, Y., & Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution*, 52(47), 47. <https://doi.org/10.1186/s12711-020-00567-1>
- Massender, E., Brito, L. F., Maignel, L., Oliveira, H. R., Jafarikia, M., Baes, C. F., Sullivan, B., & Schenkel, F. S. (2022). Single- and multiple-breed genomic evaluations for conformation traits in Canadian alpine and Saanen dairy goats. *Journal of Dairy Science*, 105(7), 5985–6000. <https://doi.org/10.3168/jds.2021-21713>
- Mazinani, M., & Rude, B. (2020). Population, world production and quality of sheep and goat products. *American Journal of Animal and Veterinary Sciences*, 15(4), 291–299. <https://doi.org/10.3844/ajavsp.2020.291.299>
- McMillan, A. J., & Swan, A. A. (2017). Weighting of genomic and pedigree relationships in single step evaluation of carcass traits in Australian sheep. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics*, 22(1), 557–560.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>

- Meyer, K., Tier, B., & Swan, A. (2018). Estimates of genetic trend for single-step genomic evaluations. *Genetics Selection Evolution*, 50(39), 1–11. <https://doi.org/10.1186/s12711-018-0410-1>
- Misztal, I., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A., & Vitezica, Z. (2018). *Manual for BLUPF90 family programs*. University of Georgia.
- Moghaddar, N., Khansef, M., Van Der Werf, J. H. J., Bolormaa, S., Duijvesteijn, N., Clark, S. A., Swan, A. A., Daetwyler, H. D., & MacLeod, I. M. (2019). Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. *Genetics Selection Evolution*, 51(72), 1–14. <https://doi.org/10.1186/s12711-019-0514-2>
- Mrode, R., Tarekegn, G. M., Mwacharo, J. M., & Djikeng, A. (2018). Invited review: Genomic selection for small ruminants in developed countries: How applicable for the rest of the world? *Animal*, 12(7), 1333–1340. <https://doi.org/10.1017/S1751731117003688>
- Mucha, A., Wierzbicki, H., Kamiński, S., Oleński, K., & Hering, D. (2019). High-frequency marker haplotypes in the genomic selection of dairy cattle. *Journal of Applied Genetics*, 60, 179–186. <https://doi.org/10.1007/s13553-019-00489-9>
- Notter, D. R. (1998). The U.S. National Sheep Improvement Program: Across-flock genetic evaluations and new trait development. *Journal of Animal Science*, 76(9), 2324–2330. <https://doi.org/10.2527/1998.7692324x>
- Oliveira, H. R., McEwan, J. C., Jakobsen, J. H., Blichfeldt, T., Meuwissen, T. H. E., Pickering, N. K., Clarke, S. M., & Brito, L. F. (2021). Across-country genomic predictions in Norwegian and New Zealand composite sheep populations with similar development history. *Journal of Animal Breeding and Genetics*, 139(1), 1–12. <https://doi.org/10.1111/jbg.12642>
- Park, L. (2011). Effective population size of current human population. *Genetics Research*, 93(2), 105–114. <https://doi.org/10.1017/S0016672310000558>
- Pattaro, C., Ruczinski, I., Fallin, D. M., & Parmigiani, G. (2008). Haplotype block partitioning as a tool for dimensionality reduction in SNP association studies. *BMC Genomics*, 9, 405. <https://doi.org/10.1186/1471-2164-9-405>
- Prakapenka, D., Wang, C., Liang, Z., Bian, C., Tan, C., & Da, Y. (2020). GVCHAP: A computing pipeline for genomic prediction and variance component estimation using haplotypes and SNP markers. *Frontiers in Genetics*, 11(282), 1–10. <https://doi.org/10.3389/fgene.2020.00282>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15(478), 478. <https://doi.org/10.1186/1471-2164-15-478>
- Stachowicz, K., Brito, L. F., Oliveira, H. R., Miller, S. P., & Schenkel, F. S. (2018). Assessing genetic diversity of various Canadian sheep breeds through pedigree analyses. *Canadian Journal of Animal Science*, 98(4), 741–749. <https://doi.org/10.1139/cjas-2017-0187>
- Taliun, D., Gamper, J., Leser, U., & Pattaro, C. (2016). Fast sampling-based whole-genome haplotype block recognition. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(2), 315–325. <https://doi.org/10.1109/TCBB.2015.2456897>
- Taliun, D., Gamper, J., & Pattaro, C. (2014). Efficient haplotype block recognition of very long and dense genetic sequences. *BMC Bioinformatics*, 51(10), 1–18. <https://doi.org/10.1186/1471-2105-15-10>
- Teissier, M., Larroque, H., Brito, L. F., Rupp, R., Schenkel, F. S., & Robert-Granié, C. (2020). Genomic predictions based on haplotypes fitted as pseudoSNPs for milk production and udder type traits and somatic cell score in French dairy goats. *Journal of Dairy Science*, 103(12), 11559–11573. <https://doi.org/10.3168/jds.2020-18662>
- Thorne, J. W., Bowdridge, S. A., Murdoch, B. M., & Redden, R. R. (2022). Response of Rambouillet lambs to an artificial gastrointestinal nematode infection. *Animals*, 12, 1199. <https://doi.org/10.3390/ani12091199>
- Thorne, J. W., Murdoch, B. M., Freking, B. A., Redden, R. R., Murphy, T. W., Taylor, J. B., & Blackburn, H. D. (2021). Evolution of the sheep industry and genetic research in the United States: Opportunities for convergence in the twenty-first century. *Animal Genetics*, 52(4), 395–408. <https://doi.org/10.1111/age.13067>
- Van Vleck, L. D. (1993). Variance of prediction error with mixed model equations when relationships are ignored. *Theoretical and Applied Genetics*, 85(5), 545–549. <https://doi.org/10.1007/BF00220912>
- Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Ventura, R. V., Miller, S. P., Dodds, K. G., Auvray, B., Lee, M., Bixley, M., Clark, S. M., & McEwan, J. C. (2016). Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. *Genetics Selection Evolution*, 48(71), 71. <https://doi.org/10.1186/s12711-016-0244-7>
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, 94(2), 73–83. <https://doi.org/10.1017/S0016672312000274>
- Wang, L., Janss, L. L., Madsen, P., Henshall, J., Huang, C.-H., Marois, D., Alemu, S., Sørensen, A. C., & Jensen, J. (2020). Effect of genomic selection and genotyping strategy on estimation of variance components in animal models using different relationship matrices. *Genetics Selection Evolution*, 52(31), 31. <https://doi.org/10.1186/s12711-020-00550-w>
- Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., & Jin, L. (2002). Distribution of recombination crossovers and the origin of haplotype blocks: The interplay of population history, recombination, and mutation. *American Journal of Human Genetics*, 71(5), 1227–1234. <https://doi.org/10.1086/344398>
- Wellmann, R. (2019). Optimum contribution selection for animal breeding and conservation: The R package optiSel. *BMC Bioinformatics*, 20(25), 25. <https://doi.org/10.1186/s12859-018-2450-5>
- Weng, Z., Wolc, A., Su, H., Fernando, R. L., Dekkers, J. C. M., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., & Garrick, D. J. (2019). Identification of recombination hotspots and quantitative trait loci for recombination rate in layer chickens. *Journal*

- of *Animal Science and Biotechnology*, 10(20), 20. <https://doi.org/10.1186/s40104-019-0332-y>
- Weng, Z.-Q., Saatchi, M., Schnabel, R. D., Taylor, J. F., & Garrick, D. J. (2014). Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genetics Selection Evolution*, 46(34), 1–14. <https://doi.org/10.1186/1297-9686-46-34>
- Won, S., Park, J.-E., Son, J.-H., Lee, S.-H., Park, B. H., Park, M., Park, W.-C., Chai, H.-H., Kim, H., Lee, J., & Lim, D. (2020). Genomic prediction accuracy using haplotypes defined by size and hierarchical clustering based on linkage disequilibrium. *Frontiers in Genetics*, 11(134), 1–9. <https://doi.org/10.3389/fgene.2020.00134>
- Xu, L., Gao, N., Wang, Z., Xu, L., Liu, Y., Chen, Y., Xu, L., Gao, X., Zhang, L., Gao, H., Zhu, B., & Li, J. (2020). Incorporating genome annotation into genomic prediction for carcass traits in Chinese Simmental beef cattle. *Frontiers in Genetics*, 11(481), 1–13. <https://doi.org/10.3389/fgene.2020.00481>
- Yu, H., & Morota, G. (2021). GCA: An R package for genetic connectedness analysis using pedigree and genomic data. *BMC Genomics*, 22, 119. <https://doi.org/10.1186/s12864-021-07414-7>
- Zhang, W., Dai, X., Wang, Q., Xu, S., & Zhao, P. X. (2016). PEPIS: A pipeline for estimating Epistatic effects in quantitative

trait locus mapping and genome-wide association studies. *PLoS Computational Biology*, 12(5), e1004925. <https://doi.org/10.1371/journal.pcbi.1004925>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Araujo, A. C., Carneiro, P. L. S., Oliveira, H. R., Lewis, R. M., & Brito, L. F. (2023). SNP- and haplotype-based single-step genomic predictions for body weight, wool, and reproductive traits in North American Rambouillet sheep. *Journal of Animal Breeding and Genetics*, 140, 216–234. <https://doi.org/10.1111/jbg.12748>