Genomic predictions based on haplotypes fitted as pseudo-SNP for milk production and udder type traits and SCS in French dairy goats

Marc Teissier, Hélène Larroque, Luiz F. Brito, Rachel Rupp, Flavio S Schenkel, Christèle Robert-Granié

J. Dairy Sci. 103
https://doi.org/10.3168/jds.2020-18662

Genomic predictions based on haplotypes fitted as pseudo-SNP for milk production and udder type traits and SCS in French dairy goats

Marc Teissier,[1]* Hélène Larroque,[1] Luiz F. Brito,[2,3] Rachel Rupp,[1] Flavio S. Schenkel,[2] and Christèle Robert-Granié[1]
[1]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326 Castanet-Tolosan, France
[2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
[3]Department of Animal Sciences, Purdue University, West Lafayette, IN 47907

ABSTRACT

The development of statistical methods aiming to improve the accuracy of genomic predictions is of utmost value for dairy goat breeding programs. In this context, the use of haplotypes, instead of individual SNP, could improve the accuracy of genomic predictions by better capturing the effect of causal variants, instead of relying solely on linkage disequilibrium with individual SNP. Haplotypes can be included in genomic evaluation models in various ways, such as fitting them as pseudo-SNP (i.e., haplotypes converted into biallelic SNP format). This can be easily incorporated in the software already available for single-step genomic predictions (ssGBLUP). Therefore, the aim of this study was to compare the predictive performances of ssGBLUP and weighted ssGBLUP (WssGBLUP) based on individual SNP or on haplotypes fitted as pseudo-SNP. Performance was compared in terms of accuracy, bias, and weights for SNP versus pseudo-SNP. Genomic predictions were performed on 5 milk production traits, 5 udder type traits, and somatic cell score (SCS). The training population was formed by 307 Alpine and 247 Saanen progeny-tested bucks, genotyped using the Illumina Goat SNP50 BeadChip (Illumina, San Diego, CA). The validation population included 205 Alpine and 146 Saanen young bucks. The accuracy of genomic predictions was evaluated in the validation population as the Pearson correlation between genomic estimated breeding values (GEBV), predicted based on various methods, and daughter deviation (DD) based on the official genetic evaluation of January 2016. Haplotype-based models were shown to improve the performance of genomic predictions for some traits. Gains in accuracy of up to +19% (0.310 to 0.368 for fat yield) in Alpine and up to +3% (0.361 to 0.373 for udder shape) in Saanen were observed with ssGBLUP. The ssGBLUP accuracies averaged across all traits and methods were equal to 0.467 (SNP) versus 0.471 (pseudo-SNP) in Alpine and 0.528 (SNP) versus 0.523 (pseudo-SNP) in Saanen. With WssGBLUP, gains in accuracy of up to 24% (0.298 to 0.370 for fat yield) in Alpine and 14% (0.431 to 0.490 for SCS) in Saanen were observed with WssGBLUP. Accuracies of WssGBLUP averaged across all traits and methods were equal to 0.455 (SNP and pseudo-SNP) in Alpine and 0.542 (SNP) versus 0.528 (pseudo-SNP) in Saanen. The average (±SD) slope of the regression of DD on GEBV for the validation animals, across all breeds, traits and scenarios, were equal to 0.82 ± 0.20 (SNP) and 0.83 ± 0.18 (pseudo-SNP) for ssGBLUP and 0.67 ± 0.16 (SNP) and 0.65 ± 0.16 (pseudo-SNP) for WssGBLUP, which suggest that haplotype-based models and ssGBLUP$_{SNP}$ were similarly biased. However, WssGBLUP was more biased than ssGBLUP, and its gains in accuracies were limited to milk production traits. Despite the fact that genomic predictions based on haplotypes require additional steps and time, the observed gains in GEBV predictive performance indicate that haplotype-based methods could be recommended for some traits.

Key words: genomic selection, haplotype-based models, individual SNP-based models, ssGBLUP, weighted ssGBLUP

INTRODUCTION

The recent adoption of genomic selection in dairy goats (Rupp et al., 2016) has nurtured the development of sophisticated genomic evaluation methods. Genomic evaluation aims to accurately identify the breeding candidates with highest genetic merit at an earlier age and therefore increase genetic progress for economically important traits. A class of genomic evaluation methods that uses large-scale genomic information to estimate genetic relationships between pairs of animals includes the genomic best linear unbiased prediction (**GBLUP**;

Received April 6, 2020.
Accepted July 27, 2020.
*Corresponding author: marc.teissier@inrae.fr

VanRaden, 2008) as well as its application in a 2-step GBLUP (e.g., Ricard et al., 2013; Edel et al., 2017) and single-step GBLUP (**ssGBLUP**; Legarra et al., 2009; Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). These methods are at least as accurate as the pedigree-based BLUP (Calus et al., 2014; Piccoli et al., 2020). However, a limitation of these GBLUP methods is that they assume that all SNP effects come from the same distribution, making no distinction between polygenic traits and traits under the influence of major genes (Legarra et al., 2009), which might not be true for complex traits with important QTL. To overcome this limitation, the weighted ssGBLUP (**WssGBLUP**) method was proposed to perform genome-wide association studies and genomic evaluations more accurately when the assumption of a single distribution is not met (Wang et al., 2012; Zhang et al., 2016). In brief, WssGBLUP allocates greater weights to SNP that are in high linkage disequilibrium (**LD**) with a causal mutation or associated with QTL with relatively large effect (Wang et al., 2012). Zhang et al. (2016) proposed alternative WssGBLUP approaches to use common weights for consecutive SNP and create the weighted genomic relationship matrix. The common weight for a defined genomic window is calculated as the sum of all SNP weights in the window (WssGBLUP$_{\text{Sum}}$) or as the maximum weight in the window (WssGBLUP$_{\text{Max}}$; Teissier et al., 2018, 2019).

Teissier et al. (2019) evaluated the accuracy of genomic predictions from ssGBLUP and WssGBLUP and its alternatives (WssGBLUP$_{\text{Sum}}$ and WssGBLUP$_{\text{Max}}$) on all traits under artificial selection in French dairy goats (Alpine and Saanen breeds). These studies showed that accuracies were up to 14% greater when using WssGBLUP and its alternatives compared with ssGBLUP for traits with a known major gene or QTL. For instance, this improvement was observed for protein content, which is affected by the α_{S1}-*casein* gene, located on *Capra hircus* chromosome 6 (**CHI 6**), which is segregating in both Alpine and Saanen goats. Milk, fat, and protein yields, udder floor position, and rear udder attachment were also improved for the Saanen breed, possibly due to a large QTL positioned on CHI 19. For the other traits, with a more polygenic genetic background, the accuracies with WssGBLUP and its alternatives were similar or slightly lower (i.e., from 5 to 0% lower) compared with ssGBLUP. These findings indicate that WssGBLUP and its alternatives are suitable methods to consider the presence of major genes or QTL in genomic evaluations.

Another promising strategy to improve the accuracy of genomic predictions is the use of haplotypes (Calus et al., 2008; Cuyabano et al., 2014; Feitosa et al., 2020). A haplotype can be defined as a group of nearby SNP on the same homologous chromosome, which are frequently inherited together (International HapMap Consortium, 2005). The use of haplotypes in genomic evaluations has some advantages compared with fitting individual SNP. Haplotypes are more informative than SNP to describe recent identical-by-descendent relationships, and they may also capture LD with multiallelic QTL better than individual SNP, which are often biallelic (Meuwissen et al., 2014). Additionally, long haplotypes might be better to differentiate identical-by-descendent and identical-by-state, as long shared haplotypes are likely to be inherited from more recent common ancestors (Broman and Weber, 1999). The SNP present on a SNP chip panel are often chosen to have moderate to high minor allele frequency. Therefore, these SNP are usually old mutations, as new ones have low frequency when they first emerge (Meuwissen et al., 2014), and, consequently, individual SNP tend to be less efficient than haplotypes to trace new mutations (Meuwissen et al., 2014).

In practice, the performance of genomic predictions based on haplotypes vary across traits and species. For instance, some studies reported no improvement in accuracy of genomic predictions based on haplotypes (Hickey et al., 2013; Meuwissen et al., 2014; Uemoto et al., 2017), whereas others showed improved accuracies when fitting haplotypes (Jónás et al., 2016; Hess et al., 2017; Karimi et al., 2018). Haplotypes can be defined by grouping together consecutive SNP (Hickey et al., 2013; Ferdosi et al., 2016), or using LD information to construct haploblocks (Cuyabano et al., 2014). Jónás et al. (2016) proposed another approach, using preselection of SNP and optimization of allele frequencies to construct haplotypes. They compared the accuracy of genomic evaluation with haplotypes constructed around QTL (selected based on the BayesCpi method) and observed an increase in genomic EBV (**GEBV**) accuracy by 0.7 to 0.9 percentage points for 5 milk production traits. A promising alternative to integrate haplotypes into genomic prediction models is to convert them into pseudo-SNP (Karimi et al., 2018; Feitosa et al., 2020), which can be easily implemented in commercial ssGBLUP software when constructing the genomic relationship matrix (**G**). In this context, the aim of the present study was to investigate alternative single-step genomic prediction methods using haplotypes fitted as pseudo-SNP compared with individual SNP markers. Accuracy and bias (slope of regression) of genomic evaluations from ssGBLUP and WssGBLUP, when fitting SNP or pseudo-SNP, were used as comparison criteria.

MATERIALS AND METHODS

Data Sets and Data Quality Control

The data sets used were provided by the French National Milk Recording System (Jouy-en-Josas, France) and were from the official genetic evaluation (Larroque et al., 2011) of January 2016. They contained phenotypes, pedigrees, genotypes (Illumina Goat SNP50 BeadChip; Illumina, San Diego, CA), and environmental effects for the Alpine and Saanen breeds. No animal handling and ethical committee approval was needed, as all the data sets were obtained from pre-existing databases.

Five milk production traits, 5 udder type traits, and SCS were considered in this study. The milk production traits were milk, fat, and protein yields (**MY**, **FY** and **PY**, respectively; in kg) and fat and protein content (**FC** and **PC**, in g/kg of milk). The numbers of phenotypes were approximately 3 and 4 million for the Saanen and Alpine breed, respectively (Table 1). The udder type traits, scored from 1 to 9, were udder floor position (**UFP**), rear udder attachment (**RUA**), udder shape (**US**), teat angle (**TA**), and fore udder (**FU**). For udder type traits, there were approximately 150,000 and 100,000 records for the Alpine and Saanen breeds, respectively (Table 1). The smaller number of records for the udder type traits is due to a single measurement being performed in each animal's lifetime and a shorter period of recording. This study also included 1.3 and 1.0 million records for SCS (calculated as log-transformed SCC) for Alpine and Saanen, respectively.

The pedigree file contained animals born between 1936 and 2012. It included 1,446,296 Alpine and 1,097,384 Saanen animals for the milk production traits. For udder type traits, the pedigree file included 290,656 Alpine and 206,154 Saanen. For SCS, the pedigree contained 788,576 Alpine and 648,461 Saanen individuals. Unknown parent groups were also defined, in which one group included all animals born before 1975 and then pooled groups (sires and dams) were defined every 2 yr. Sires and dams were pooled together, as few animals had only unknown dams.

A total of 2,056 Alpine and 1,349 Saanen animals born between 1980 and 2012 were genotyped with the Illumina Goat SNP50 BeadChip (50K; Tosser-Klopp et al., 2014), which contained 53,347 SNP. Quality control was applied within breed, and details of the quality control procedure can be found in Teissier et al. (2018). A total of 46,849 SNP and 1,749 Alpine (512 males and 1,237 females) and 1,206 Saanen (393 males and 813 females) remained in the genomic data set for further analyses.

Genomic Predictions Fitting Individual SNP

ssGBLUP Method. The ssGBLUP method uses simultaneously all phenotypes, pedigree records, and genotypes to estimate GEBV for all animals included in the analyses (Legarra et al., 2009). Genomic evaluations in this study were based on 2 different models: one for milk production traits and SCS and another for udder type traits. Analyses were performed within breed. For milk production traits and SCS, the model used was as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}, \qquad \text{[Model 1]}$$

where $\mathbf{y}$ is a vector of phenotypes (MY, FY, PY, FC, PC, or SCS), and $\boldsymbol{\beta}$ is the vector of fixed effects, which included herd by year of the phenotypic measurement by parity, birth year by year of the phenotypic measurement by region, birth month by year of the phenotypic measurement by region, and length of dry period (assigned to 10 classes within each parity) by year of the phenotypic measurement by region. There were

Table 1. Descriptive statistics and heritability estimates (h^2) for the 11 traits of Alpine and Saanen goat breeds included in this study[1]

	Alpine			Saanen		
Trait	Number of records	Mean (±SD)	h^2	Number of records	Mean (±SD)	h^2
Milk yield	3,844,314	802.12 ± 247.37	0.31	2,923,531	823.08 ± 259.40	0.26
Fat yield	3,742,129	28.4 ± 9.69	0.28	2,887,051	27.44 ± 9.51	0.25
Protein yield	3,844,071	24.36 ± 7.87	0.31	2,923,419	24.32 ± 7.74	0.25
Fat content	3,742,129	35.33 ± 5.18	0.48	2,887,051	33.39 ± 4.96	0.51
Protein content	3,844,071	30.42 ± 3.29	0.60	2,923,419	29.68 ± 2.83	0.56
Teat angle	150,676	3.63 ± 0.90	0.42	102,967	4.05 ± 0.85	0.45
Udder floor position	150,676	6.37 ± 1.05	0.51	102,967	6.16 ± 1.15	0.57
Rear udder attachment	150,676	4.57 ± 1.45	0.47	102,967	4.96 ± 1.62	0.52
Fore udder	150,676	3.19 ± 1.00	0.44	102,967	3.38 ± 1.16	0.42
Udder shape	150,676	5.76 ± 1.39	0.40	102,967	6.22 ± 1.33	0.47
SCS	1,262,187	8.52 ± 1.38	0.20	1,031,450	8.71 ± 1.31	0.16

[1]Heritability estimated in Carillier et al., 2014.

8,250 herds, 32 birth years (from 1980 to 2012), and 3 parities (1, 2, or ≥3). France is divided into 4 geographical regions by the goat breeding management program, and these regions were also fitted in the model. The same fixed effects were used for milk production traits and SCS. $\mathbf{u}$ is a vector of GEBV assumed to be normally distributed $N\left(0, \mathbf{H}\sigma_u^2\right)$, where $\mathbf{H}$ represent the hybrid relationship matrix (Legarra et al., 2009), $\mathbf{p}$ is a vector of random permanent environmental effects assumed to be normally distributed $N\left(0, \mathbf{I}\sigma_p^2\right)$, $\mathbf{I}$ is the identity matrix, and $\mathbf{e}$ is a vector of random residuals normally distributed $N\left(0, \mathbf{I}\sigma_e^2\right)$. $\mathbf{X}$ is the incidence matrix relating phenotypes to the fixed effects ($\boldsymbol{\beta}$); $\mathbf{Z}$ is the incidence matrix relating phenotypes to the GEBV ($\mathbf{u}$); and $\mathbf{W}$ is the incidence matrix relating the phenotypes to the permanent environmental effects ($\mathbf{p}$). For udder type traits, the following model was used:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \qquad \text{[Model 2]}$$

where $\mathbf{y}$, $\mathbf{u}$, and $\mathbf{e}$ were the same vectors as described for [Model 1], and $\boldsymbol{\beta}$ is the vector of fixed effects, which included herd by year by parity, age at scoring by year, and lactation stage at scoring by year. The year effect had 32 levels (from 1980 to 2012), and parity had 2 levels (1 and 2).

The inverse of the $\mathbf{H}$ matrix can be obtained thus:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}$ is the numerator relationship matrix estimated based on the pedigree information, $\mathbf{A}_{22}$ is the $\mathbf{A}$ matrix for genotyped animals, and $\mathbf{G}$ is the genomic relationship matrix constructed as in the first VanRaden (2008) method:

$$\mathbf{G} = \frac{\mathbf{M}'\mathbf{M}}{2\sum_{i=1}^{m} p_i\left(1 - p_i\right)},$$

where m is the total number of SNP, p_i is the observed allele frequency at locus i, and $\mathbf{M}$ is a centered matrix of SNP genotypes with elements $\mathbf{M}_{ij} = \mathbf{P}_{ij} - 2(p_j - 0.5)$, where $\mathbf{P}$ is a matrix of genotypes (coded −1, 0, 1). In this study, results from ssGBLUP based on individual SNP (using $\mathbf{G}$) will be termed as $\mathbf{ssGBLUP_{SNP}}$. The analyses were performed using blup90iod2 software (Misztal et al., 2016).

Weighted ssGBLUP Method. The WssGBLUP method allocates SNP weights according to their effect on the trait to estimate the genetic relationship between each pair of animals. This method, proposed by Wang et al. (2012) and based on a model similar to ssGBLUP, gives different weights for each SNP and enables better fitting of major genes or QTL with a relatively large effect using a weighted $\mathbf{G}$ matrix ($\mathbf{G}^*$). $\mathbf{G}^*$ can be defined as follows:

$$\mathbf{G}^* = \frac{\mathbf{M}'\mathbf{D}\mathbf{M}}{2\sum_{i=1}^{m} p_i\left(1 - p_i\right)},$$

where $\mathbf{A}_{22}$, $\mathbf{M}$, p_i, and m are the same as in $\mathbf{G}$, and $\mathbf{D}$ is a diagonal matrix of size $m \times m$, where each element of the diagonal corresponds to a SNP weight. Weights of SNP were calculated based on the GEBV estimated using ssGBLUP. The WssGBLUP approach is based on an iterative algorithm with different steps: (1) run ssGBLUP with the $\mathbf{G}^*$ matrix (at iteration 1, the SNP weights in the $\mathbf{D}$ matrix are equal to 1 and equivalent to ssGBLUP); (2) estimate SNP effects from solutions of GEBV in the previous step; (3) estimate variances of the effect of each SNP; (4) normalize the vector of variances of SNP effects to get the SNP weights (this normalization process ensures that the sum of the variances remain constant and equal to the number of SNP); (5) use SNP weights to construct the $\mathbf{D}$ matrix; (6) loop to step 1. Previous studies have shown that the second iteration of the WssGBLUP yields the most accurate GEBV (Wang et al., 2012; Teissier et al., 2018), and therefore results will be presented only for this scenario. Results from WssGBLUP based on individual SNP will be termed $\mathbf{WssGBLUP_{SNP}}$. SNP effects, weights, and GEBV were estimated using blup90iod2 software (Misztal et al., 2016).

Haplotypic Genomic Predictions

In this study, haplotypes were constructed based on 2 different methods: either by considering a fixed number of adjacent SNP along the chromosome, called the distinct windows (**DW**) method (Ferdosi et al., 2016), or by using a fixed LD threshold between each pair of SNP, defined here as the $\mathbf{LD_{hap}}$ method (Cuyabano et al., 2014). The construction of haplotypes requires phased genotypes; thus, parental haplotypes were reconstructed using the FImpute software (Sargolzaei et al., 2014). The phasing step was performed within breed using all available genotypes. Figure 1 illustrates the haplotype construction based on the DW and $\mathrm{LD_{hap}}$ methods (Cuyabano et al., 2014; Ferdosi et al., 2016),

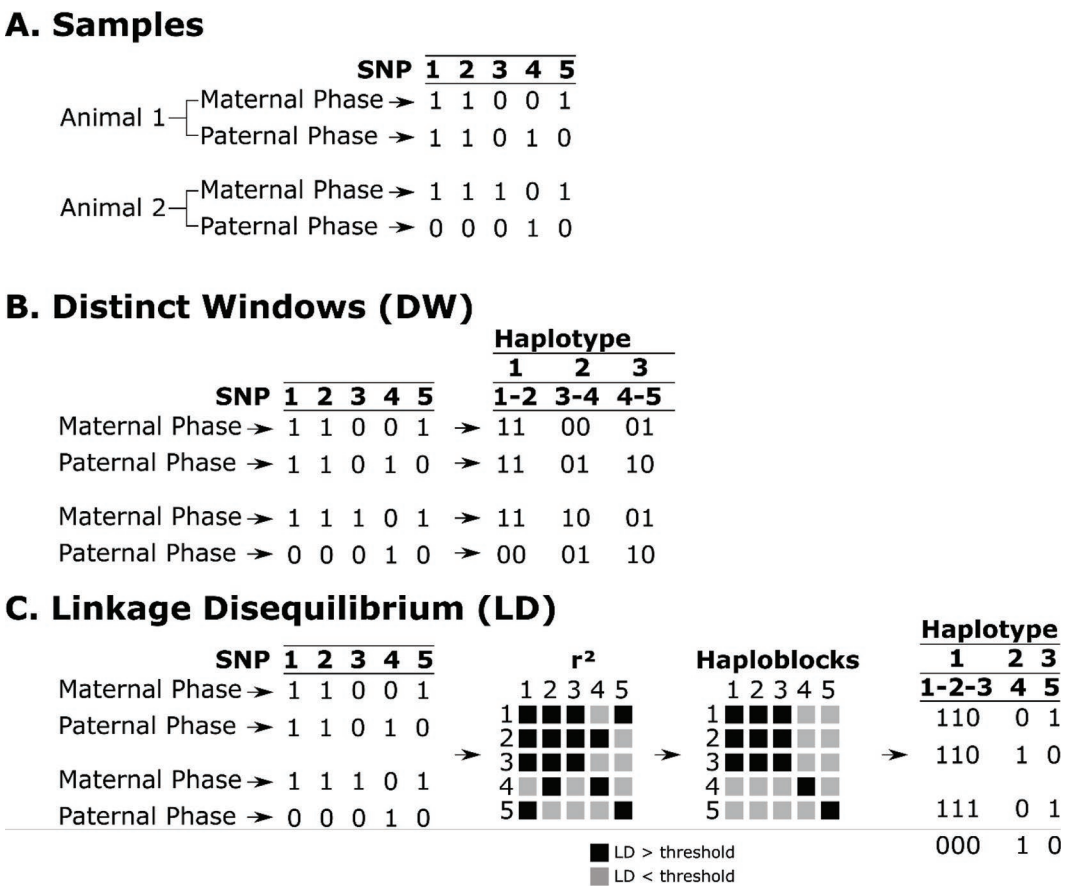

Figure 1. Construction of haplotypes using the distinct windows (DW) or linkage disequilibrium (LD) methods. Initially, genotypes are phased (A). In DW (B), the size of the window required to create the haplotypes needs to be defined (here, 2 SNP). In LD (C), LD between SNP needs to be estimated before construction of the haplotypes.

using an example of 2 animals genotyped for 5 SNP and having known parental haplotypes.

Haplotype Construction

Distinct Windows Method (DW). The haplotypes from the DW method were defined by considering a fixed number of adjacent SNP along the chromosome (Hickey et al., 2013; Ferdosi et al., 2016), as shown in Figure 1B. For the last haplotype of the chromosome, if the number of adjacent SNP was shorter than those predefined, SNP from the previous fragment were used to ensure that all haplotypes were of equal size. For instance, in the example illustrated in Figure 1B, SNP 4 was present in haplotypes 2 and 3. The haplotype lengths investigated in this study were 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 SNP.

Linkage Disequilibrium Method (LD_{hap}). The haplotypes based on the LD_{hap} method were constructed as suggested by Cuyabano et al. (2014) and are called haploblocks. First, LD was computed between all pairs of SNP using PLINK software (Purcell et al., 2007) and the r^2 metric (Rogers and Huff, 2009). This measure ranges from 0 (no LD) to 1 (complete LD between 2 SNP):

$$r^2 = \frac{\left[\mathrm{cov}\left(g_i, g_j\right)\right]^2}{\mathrm{var}\left(g_i\right) \times \mathrm{var}\left(g_j\right)},$$

where g_i and g_j are the genotypes (coded as 0, 1, or 2) for SNP i and j. A haploblock was defined as a group of SNP in which the LD between each pair of SNP was equal to or higher than a fixed threshold. In the example shown in Figure 1C, the LD between 2 SNP higher than the threshold is represented in black; otherwise it is shown in gray. In the figure, haploblocks are presented as a square where all cells are shown in black; this includes SNP 1, 2, and 3, which are grouped in the same haploblock. The LD between SNP 4 and 5 was not high enough to be considered a haploblock with 2

SNP, so 2 "haploblocks" with only one SNP were created. Thresholds LD of 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1 were evaluated.

The Haplotypic Genomic Relationship Matrix

Haplotypes were converted to pseudo-SNP and fitted in the genomic prediction models through the genomic relationship matrix. To achieve this, each allele of each haplotype was considered as a pseudo-SNP, and its number of copies was counted for each animal and phase (allele count was equal to 0, 1, or 2). The total number of haplotypes was different depending on the size of haplotype used. Figure 2 shows the results of the transformation of haplotypes to pseudo-SNP from the example illustrated in Figure 1.

Pseudo-SNP with frequency lower than 1% were filtered out from further analyses. Because haplotypes are coded as SNP, the implementation of ssGBLUP and WssGBLUP with pseudo-SNP is straightforward. Thus, pseudo-SNP were used to construct the genomic relationship matrix $\mathbf{G}_{\text{pseudo-SNP}}$ for the ssGBLUP and $\mathbf{G}^*_{\text{pseudo-SNP}}$ for the WssGBLUP:

$$\mathbf{G}_{\text{pseudo-SNP}} = \frac{\mathbf{M}'_{\text{pseudo-SNP}}\mathbf{M}_{\text{pseudo-SNP}}}{2\sum_{i=1}^{m} p_i\left(1-p_i\right)}$$

or

$$\mathbf{G}^*_{\text{pseudo-SNP}} = \frac{\mathbf{M}'_{\text{pseudo-SNP}}\mathbf{D}\mathbf{M}_{\text{pseudo-SNP}}}{2\sum_{i=1}^{m} p_i\left(1-p_i\right)},$$

where $\mathbf{M}_{\text{pseudo-SNP}}$ is a centered matrix of pseudo-SNP, constructed based on either the DW or the LD_{hap} method, and $\mathbf{M}'_{\text{pseudo-SNP}}$ is the transpose of the $\mathbf{M}_{\text{pseudo-SNP}}$ matrix. The methods using pseudo-SNP will be termed **ssGBLUP$_{\text{pseudo-SNP(DW)}}$** or **WssGBLUP$_{\text{pseudo-SNP(DW)}}$** for the DW method and **ssGBLUP$_{\text{pseudo-SNP(LD)}}$** or **WssGBLUP$_{\text{pseudo-SNP(LD)}}$** for the LD_{hap} method.

Accuracy and Bias (Inflation or Deflation) of Genomic Predictions

The genotyped animals were split into 2 subsets: a training and a validation population. The training population included 307 Alpine and 247 Saanen bucks born between 1993 and 2007, and all information on these animals (genotype, ancestral pedigree information, their progeny, and their progenies' phenotypes) was kept to estimate the GEBV for each trait. The validation set included 205 Alpine and 146 Saanen bucks born between 2008 and 2012. For these animals, progeny phenotypes were removed from the analyses, and only the genotypes and ancestor pedigree information were retained. The performance of genomic predictions was measured as the squared Pearson correlation between GEBV and daughter deviation (**DD**;

A. Distinct Windows (DW)

	Haplotype 1	Haplotype 2	Haplotype 3
SNP	1-2	3-4	4-5
Maternal Phase →	11	00	01
Paternal Phase →	11	01	10
Maternal Phase →	11	10	01
Paternal Phase →	00	01	10

Haplotype	1		2			3	
pseudo-SNP	11	00	00	01	10	01	10
Animal 1	2	0	1	1	0	1	1
Animal 2	1	1	0	1	1	1	1

B. Linkage Disequilibrium (LD)

	Haplotype 1	Haplotype 2	Haplotype 3
SNP	1-2-3	4	5
Maternal Phase →	110	0	1
Paternal Phase →	110	1	0
Maternal Phase →	111	0	1
Paternal Phase →	000	1	0

Haplotype	1			2		3	
pseudo-SNP	110	111	000	0	1	0	1
Animal 1	2	0	0	1	1	1	1
Animal 2	0	1	1	1	1	1	1

Figure 2. Construction of pseudo-SNP from haplotypes using the distinct windows (DW; A) and linkage disequilibrium (LD; B) methods. Pseudo-SNP were constructed based on the number of copies of an individual haplotype (haplotype alleles). The number of pseudo-SNP for one individual haplotype is equal to the number of alleles for this haplotype.

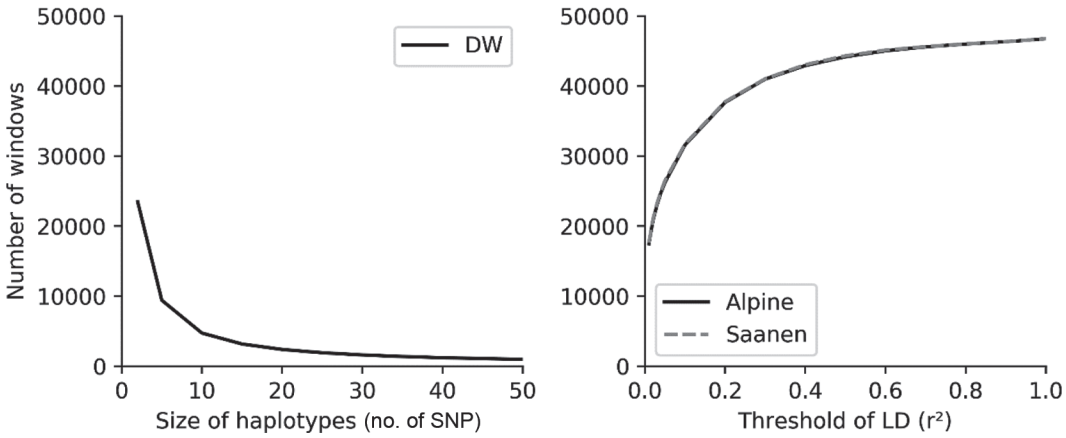

Figure 3. Number of windows according to the size of haplotype from the distinct windows (DW) method or according the threshold of linkage disequilibrium (LD) in the LD method in Alpine and Saanen goat breeds.

VanRaden and Wiggans, 1991) from the official genetic evaluation in January 2016, in the validation population. The bias (inflation) of the GEBV was assessed based on the slope of the regression of DD on GEBV for the validation animals, for which a slope of 1 indicates no inflation or deflation.

RESULTS

Number of Haplotypes and Pseudo-SNP Based on the DW and LD_{hap} Methods

We first investigated the number of windows and pseudo-SNP that were created based on the DW and LD_{hap} methods. Figure 3 presents the number of windows according to the size of haplotypes (DW) or the LD threshold (LD_{hap}). The numbers of windows with the DW method in both breeds were identical, as exactly the same SNP were retained after quality control (46,849 SNP). For DW, the number of windows decreased when size of haplotypes increased; that is, it ranged from 23,429 genomic windows with a size of 2 SNP to 937 genomic windows with a size of 50 SNP, because rare haplotypes were removed. With LD_{hap}, SNP can also be obtained, so this count includes both haplotypes and single SNP. For the LD_{hap} method, the number of genomic windows increased with a higher LD threshold. The number of windows was equal to 13,635 (+3,810 single SNP) for Alpine and 13,699 (+3,895 single SNP) for Saanen, with an LD threshold equal to 0.01. The number of genomic windows reached 482 (+45,817 single SNP) in Alpine and 481 (+45,808 single SNP) in Saanen, for an LD threshold equal to 0.9 for Alpine and Saanen. The numbers of windows in Alpine and Saanen were almost identical, with an average difference of 110 windows across the LD thresholds.

In the LD_{hap} method, a proportion of the windows was formed by individual SNP. For instance, about 10%, 50%, and 90% of the genomic windows were individual SNP for a threshold of LD of 0.01, 0.1, and 0.5, respectively (results not shown).

Figure 4 presents the number of pseudo-SNP according to the DW or LD_{hap} method. For DW, the number of pseudo-SNP increased with the size of haplotypes between 2 and 5 SNP. A maximum number of pseudo-SNP was reached for 5 SNP, with 118,151 and 117,566 pseudo-SNP in Alpine and Saanen, respectively. Subsequently, the number of pseudo-SNP decreased to 22,029 in Alpine and 19,674 in Saanen for haplotypes of size equal to 50 SNP. In Alpine and Saanen, 96% of pseudo-SNP remained after filtering based on their allele frequency for haplotypes of 2 SNP; this proportion decreased with the size of haplotypes and reached only 6% for haplotypes containing 50 SNP. In the LD_{hap} method, the highest number of pseudo-SNP was observed for an LD threshold of 0.01 (95,233 pseudo-SNP in Alpine and 94,980 pseudo-SNP in Saanen). Thereafter, the number of pseudo-SNP decreased rapidly and was lower than 50,000 (in both Alpine and Saanen breeds) for an LD threshold equal to 0.5. Finally, the number of pseudo-SNP reached 46,849 in both breeds, with an LD threshold equal to 1 (i.e., only individual SNP remained in the analyses).

Pseudo-SNP Weights

We investigated weights for pseudo-SNP in Alpine and Saanen for all traits. A well-known gene (α_{S1}-*casein*) is associated with PC in Alpine and Saanen goats on CHI 6. Therefore, we used PC to observe the effects of DW and LD_{hap} on pseudo-SNP weights for SNP close to the α_{S1}-*casein* gene. Figure 5 presents weights for

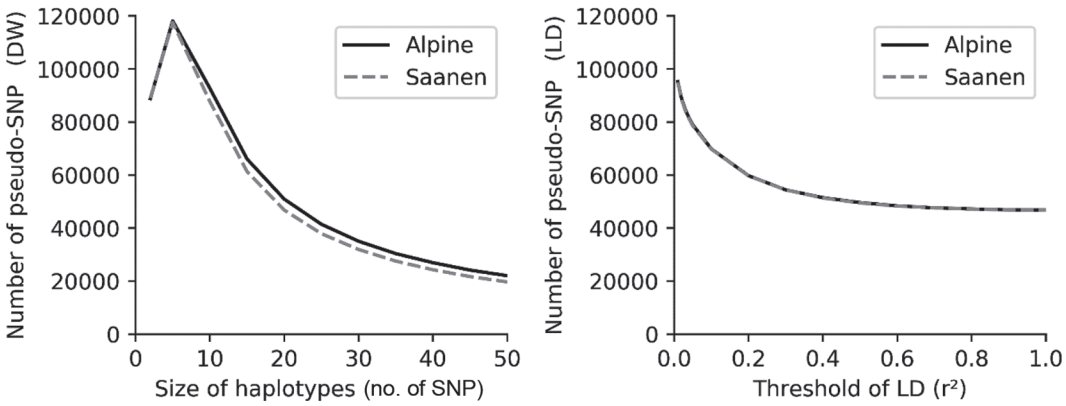

Figure 4. Number of pseudo-SNP used in genomic predictions, after filtering out alleles with a frequency lower than 0.01 for the distinct windows (DW) and linkage disequilibrium (LD) methods in Alpine and Saanen goat breeds.

pseudo-SNP located on CHI 6 for PC in the Alpine and Saanen breeds according to the different sizes of haplotypes (DW) used in this study. Large weights for pseudo-SNP were located at the extremities of CHI 6 for both breeds. In Alpine, the maximum weight for pseudo-SNP ranged from 54 (2-SNP haplotypes) to 454 (50-SNP haplotypes). For 2-SNP haplotypes, the sum of weights of the 1% pseudo-SNP with the highest weights explained 15% of the sum of weights of all CHI 6. It reached 64% for haplotypes with 50 SNP. In Saanen, the maximum weights were equal to 92, 410, 964, 661, 909, and 728 for haplotypes with 2, 5, 10, 15, 20, and 25 SNP, respectively. The maximum pseudo-SNP weights for longer haplotypes were equal to 191 in Saanen, averaged across all traits. For haplotypes with 5, 10, 15, 20, and 25 SNP, the sum of weights of the 1% largest pseudo-SNP explained approximately 40% of the sum of weights of all CHI 6 haplotypes; for the other chromosomes, it reached 35% on average.

Figure 6 presents the pseudo-SNP weights on CHI 6 for PC according to the LD threshold in the Alpine and Saanen breeds. For the DW method, important weights were observed at the extremities of CHI 6. When the LD threshold increased, the maximum weight of pseudo-SNP decreased in both breeds. In Alpine, the maximum weights were equal to 82 for an LD threshold equal to 0.01 and 35 for an LD threshold equal to 1. In the Saanen breed, the maximum weight was equal

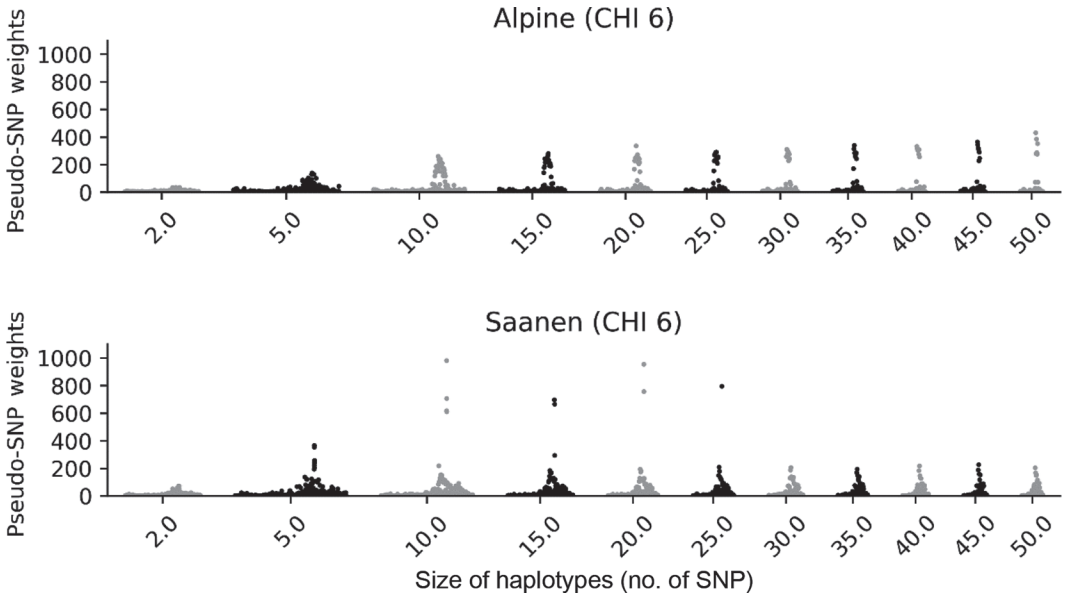

Figure 5. Pseudo-SNP weights estimation for protein content in Alpine and Saanen goat breeds for chromosome 6 (CHI 6) using the distinct windows (DW) method according to the size of haplotypes.

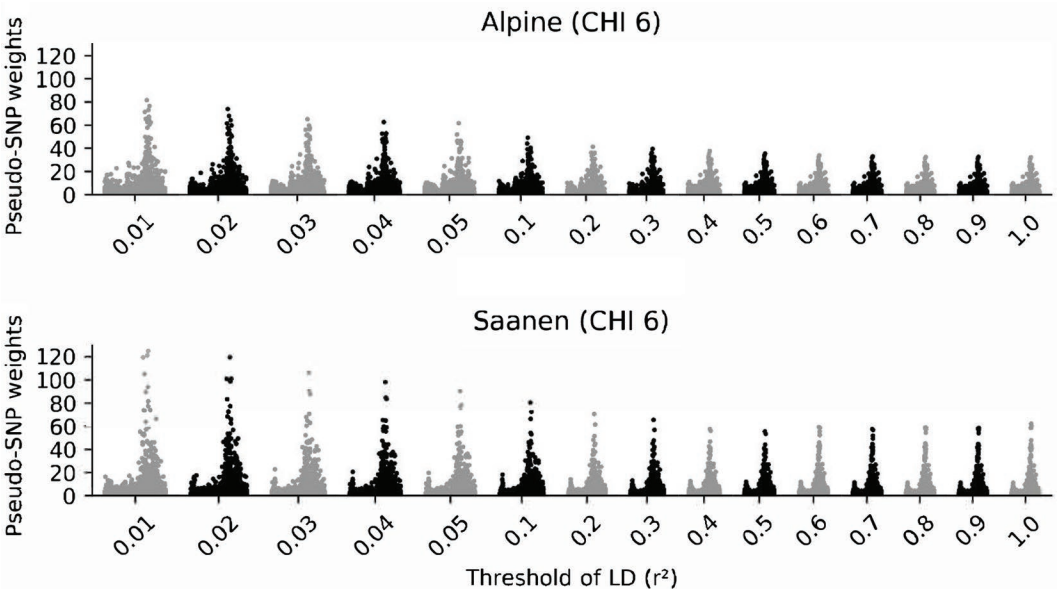

Figure 6. Pseudo-SNP weights estimation for chromosome 6 (CHI 6) for protein content in Alpine and Saanen using the linkage disequilibrium (LD) method.

to 125 for an LD threshold of 0.01 and equal to 67 for an LD threshold equal to 1. In Alpine, and with an LD threshold equal to 0.01, the sum of weights of the 1% pseudo-SNP with the highest weights explained 24% of the sum of weights of all CHI 6 haplotypes. With an LD threshold of 1, this proportion decreased to approximately 13%. The same trend was observed for the Saanen breed, in which 23% of all pseudo-SNP weights of CHI 6 was explained by the 1% pseudo-SNP with the highest weights for an LD threshold equal to 0.01. This proportion was approximately constant according to the LD threshold and was equal to 21% for an LD threshold equal to 1.

The same trend observed for PC in the Saanen breed was also observed for MY, FY, PY, FU, RUA, UFP, and SCS on CHI 19. The SNP weights on CHI 19 decreased with increasing LD threshold. A smaller peak was observed on CHI 13 for FU based on the DW method. For the other traits, no peaks (i.e., pseudo-SNP with high weights) were observed. For the Alpine breed, no peaks were detected for any trait, except for PC (results not shown).

Accuracy and Bias of Genomic Prediction

Table 2 presents the accuracies of genomic predictions with individual SNP ($ssGBLUP_{SNP}$) and haplotypes [$ssGBLUP_{pseudo\text{-}SNP(DW)}$ and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$] for each trait in the Alpine breed. Only the best and lowest accuracy estimates according to the haplotype size (DW) or threshold of LD (LD_{hap}) are presented. The accuracies for each individual trait are presented in Supplemental Tables S1 and S2 (https://doi.org/10.3168/jds.2020-18662). In general, we observed slightly higher accuracies for $ssGBLUP_{pseudo\text{-}SNP(DW)}$ or $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ compared with $ssGBLUP_{SNP}$ for MY, FY, PY, FC, FU, UFP, US, TA, and SCS, with accuracies improving by +1 to +6 percentage points for the best scenario. For other traits, slight decreases in GEBV accuracies were observed compared with $ssGBLUP_{SNP}$ (from −1 to −3 points). Accuracies were identical between $ssGBLUP_{SNP}$, $ssGBLUP_{pseudo\text{-}SNP(DW)}$, and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ for PC (0.76) and RUA (0.40). Accuracies between $ssGBLUP_{pseudo\text{-}SNP(DW)}$ and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ were generally equal between traits, but slight (nonsignificant, $P > 0.05$) differences were observed for MY (0.47 and 0.46, respectively), PY (0.32 and 0.30, respectively), and US (0.49 and 0.50, respectively). Significant ($P < 0.05$) differences were observed for FY (0.37 and 0.33, respectively). For $ssGBLUP_{pseudo\text{-}SNP(DW)}$, the highest accuracies were mainly obtained when fitting long haplotypes (25 to 50 SNP) for MY, FY, PY, FC, US, and RUA. For other traits (PC, TA, UFP, and FU), the haplotype length that yielded the highest accuracies was short and contained only 2 SNP. For LD_{hap}, the best accuracies were obtained with the threshold of 0.01 for MY, FC, UFP, US, and SCS; with a threshold of 0.02 for FY, TA, and RUA; with a threshold of 0.05 for PC; with a threshold of 0.1 for FU; and with a threshold of 0.3 for PY. For the scenario with the lowest accuracy, equal or slightly higher accuracies (0 to +1 percentage points) were observed

Table 2. Accuracy of genomic predictions for single-step genomic BLUP (ssGBLUP) based on individual SNP and pseudo-SNP with distinct windows (DW) and linkage disequilibrium (LD) methods for the Alpine breed

Trait	Scenario	SNP[1]	DW[2]	LD[3]	Window (no. of SNP) with DW	Threshold with LD
Milk yield	Best	0.45	0.47	0.46	40	0.01
	Lowest		0.46	0.45	10	[0.6, 1]
Fat yield	Best	0.31	0.37	0.33	40	0.02
	Lowest		0.32	0.31	2	[0.7, 1]
Protein yield	Best	0.30	0.32	0.30	40	0.30
	Lowest		0.30	0.30	10	(0.7, 1)
Fat content	Best	0.66	0.67	0.67	50	0.01
	Lowest		0.66	0.66	35	[0.5, 1]
Protein content	Best	0.76	0.76	0.76	2	0.05
	Lowest		0.73	0.75	50	[0.02, 0.03]
Teat angle	Best	0.42	0.43	0.43	2	0.02
	Lowest		0.40	0.42	45–50	[0.8, 0.9]
Udder floor position	Best	0.43	0.44	0.44	2	0.01
	Lowest		0.42	0.43	50	[0.4, 1]
Rear udder attachment	Best	0.40	0.40	0.40	35	0.02
	Lowest		0.39	0.40	40–45	(0.01, 0.1, 0.4, 0.8–0.9)
Fore udder	Best	0.49	0.50	0.50	2	0.10
	Lowest		0.48	0.49	50	0.03
Udder shape	Best	0.48	0.49	0.50	25	0.01
	Lowest		0.48	0.48	35	0.3
SCS	Best	0.45	0.46	0.46	5	0.01
	Lowest		0.45	0.45	40	0.8–1

[1]Genomic predictions with ssGBLUP based on individual SNP ($ssGBLUP_{SNP}$).

[2]Genomic predictions with ssGBLUP based on pseudo-SNP constructed using the DW method [$ssGBLUP_{pseudo\text{-}SNP(DW)}$].

[3]Genomic predictions with ssGBLUP based on pseudo-SNP constructed using the LD method [$ssGBLUP_{pseudo\text{-}SNP(LD)}$].

for 6 out of 11 traits with $ssGBLUP_{pseudo\text{-}SNP}$ compared with $ssGBLUP_{SNP}$ (MY, FY, PY, FC, US, SCS). For PC, TA, UFP, RUA, and FU, a decrease in accuracy of up to 3 points was observed with $ssGBLUP_{pseudo\text{-}SNP}$ compared with $ssGBLUP_{SNP}$.

Table 3 presents the accuracies of the $ssGBLUP_{SNP}$, $ssGBLUP_{pseudo\text{-}SNP(DW)}$, and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ methods for each trait in the Saanen breed. Accuracies were similar between $ssGBLUP_{SNP}$, $ssGBLUP_{pseudo\text{-}SNP(DW)}$, and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ for MY (0.49), FC (0.59), PC (0.73), FU (0.62), UFP (0.59), and RUA (0.60). Genomic prediction accuracies with $ssGBLUP_{pseudo\text{-}SNP(DW)}$ were +1 percentage point greater than $ssGBLUP_{SNP}$ for FY, PY, US, and TA. Similar results were observed with $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$. The highest improvement of accuracy was observed for SCS, with a gain of +3 percentage points with $ssGBLUP_{pseudo\text{-}SNP(DW)}$ compared with $ssGBLUP_{SNP}$. In contrast with the Alpine breed, the highest GEBV accuracies with the $ssGBLUP_{pseudo\text{-}SNP(DW)}$ method for Saanen goats were obtained when fitting short haplotypes (containing 10 SNP or less) for all traits, except PY (20 SNP), TA (15 SNP), and SCS (15 SNP). For $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$, the highest accuracies were observed with an LD threshold equal to 1 for MY, FY, PC, UFP, and RUA, where haplotypes were formed by individual SNP.

In contrast with the results for the Alpine breed, $ssGBLUP_{pseudo\text{-}SNP(DW)}$ and $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ had accuracies lower than $ssGBLUP_{SNP}$ for all the traits for Saanen. The decreases of accuracies ranged between 1 and 4 points compared with $ssGBLUP_{SNP}$. The greatest loss was for UFP between $ssGBLUP_{pseudo\text{-}SNP(DW)}$ (0.55) and $ssGBLUP_{SNP}$ (0.59).

Table 4 shows the GEBV accuracies from $WssGBLUP_{SNP}$, $WssGBLUP_{pseudo\text{-}SNP(DW)}$, and $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ for each trait and breed. To facilitate comparison with the ssGBLUP results, results of Table 4 use the same window (DW) and LD (LD_{hap}) as presented in Table 3. For the Alpine breed, genomic predictions were on average as accurate with $WssGBLUP_{pseudo\text{-}SNP(DW)}$ (0.47 ± 0.12) or $WssGBLUP_{pseudo\text{-}SNP\ (LDhap)}$ (0.47 ± 0.13) and $WssGBLUP_{SNP}$ (0.46 ± 0.14). The $WssGBLUP_{pseudo\text{-}SNP(DW)}$ and $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ were both more accurate than $WssGBLUP_{SNP}$ for FY, TA, UFP, FU, and SCS (+1 to +7 percentage points). The $WssGBLUP_{pseudo\text{-}SNP(DW)}$ were slightly more accurate than $WssGBLUP_{SNP}$ for MY and RUA (+3 to +4 percentage points), and $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ were more accurate than $WssGBLUP_{SNP}$ for PY (+1 percentage point). For the other traits, the GEBV accuracies for $WssGBLUP_{pseudo\text{-}SNP(DW)}$ and $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ were equal to or lower than the accuracies obtained

Table 3. Accuracies of genomic predictions for single-step genomic BLUP (ssGBLUP) based on individual SNP and pseudo-SNP with distinct windows (DW) and linkage disequilibrium (LD) methods for the Saanen breed

Trait	Scenario	SNP[1]	DW[2]	LD[3]	Window (no. of SNP) with DW	Threshold with LD
Milk yield	Best	0.49	0.49	0.49	5	1
	Lowest		0.47	0.47	40	0.01
Fat yield	Best	0.43	0.44	0.43	5	1
	Lowest		0.42	0.41	50	0.01
Protein yield	Best	0.45	0.46	0.45	20	0.10
	Lowest		0.44	0.44	40	0.01
Fat content	Best	0.59	0.59	0.59	5	0.02
	Lowest		0.56	0.58	50	0.01
Protein content	Best	0.73	0.73	0.73	5	1
	Lowest		0.72	0.71	45	0.01
Teat angle	Best	0.48	0.49	0.49	15	0.01
	Lowest		0.47	0.48	50	0.2
Udder floor position	Best	0.59	0.59	0.59	10	1
	Lowest		0.55	0.58	45	0.01
Rear udder attachment	Best	0.60	0.60	0.60	10	1
	Lowest		0.57	0.59	50	0.01
Fore udder	Best	0.62	0.62	0.62	2	0.10
	Lowest		0.60	0.61	50	0.02
Udder shape	Best	0.36	0.37	0.36	5	0.02
	Lowest		0.34	0.36	50	[0.04, 0.05]
SCS	Best	0.46	0.49	0.46	15	0.02
	Lowest		0.46	0.45	2	[0.02, 0.04]

[1]Genomic predictions with ssGBLUP based on individual SNP ($ssGBLUP_{SNP}$).

[2]Genomic predictions with ssGBLUP based on pseudo-SNP constructed using the DW method [$ssGBLUP_{pseudo\text{-}SNP(DW)}$].

[3]Genomic predictions with ssGBLUP based on pseudo-SNP constructed using the LD method [$ssGBLUP_{pseudo\text{-}SNP(LD)}$].

with $WssGBLUP_{SNP}$. For Saanen, the results are more method-dependent. The $WssGBLUP_{pseudo\text{-}SNP(DW)}$ slightly outperformed $WssGBLUP_{SNP}$, especially for SCS (+6 percentage points). On the other hand, the GEBV accuracies from $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ were as accurate as or less accurate than $WssGBLUP_{SNP}$ for all the traits (−1 to +1 percentage point) except for SCS (+4 percentage points).

The average slope of the regression of DD on GEBV was calculated as an indicator of GEBV bias (inflation or deflation). The slope values were averaged across all 11 traits included in this study. The regression slopes

Table 4. Accuracy of genomic evaluations based on weighted single-step genomic BLUP (WssGBLUP) with individual SNP and pseudo-SNP [using the distinct windows (DW) and linkage disequilibrium (LD) methods] in the Alpine and Saanen goat breeds[1]

	Alpine			Saanen		
Trait	SNP[2]	DW[3]	LD[4]	SNP[2]	DW[3]	LD[4]
Milk yield	0.43	0.46	0.43	0.56	0.57	0.56
Fat yield	0.30	0.37	0.33	0.48	0.49	0.48
Protein yield	0.28	0.28	0.29	0.50	0.50	0.49
Fat content	0.65	0.60	0.65	0.59	0.60	0.59
Protein content	0.77	0.77	0.77	0.77	0.75	0.77
Teat angle	0.41	0.43	0.42	0.47	0.45	0.47
Udder floor position	0.41	0.44	0.44	0.63	0.61	0.63
Rear udder attachment	0.38	0.42	0.38	0.61	0.62	0.61
Fore udder	0.48	0.49	0.49	0.59	0.61	0.60
Udder shape	0.48	0.44	0.48	0.34	0.35	0.32
SCS	0.42	0.45	0.46	0.43	0.49	0.47

[1]The highest accuracies obtained with the ssGBLUP scenario are presented (see Table 2 and Table 3 for Alpine and Saanen, respectively).

[2]Genomic predictions with WssGBLUP based on individual SNP ($WssGBLUP_{SNP}$).

[3]Genomic predictions with WssGBLUP based on pseudo-SNP constructed using the DW method [$WssGBLUP_{pseudo\text{-}SNP(DW)}$].

[4]Genomic predictions with WssGBLUP based on pseudo-SNP constructed using the LD method [$WssGBLUP_{pseudo\text{-}SNP(LD)}$].

for each trait are presented in Supplemental Tables S1 and S2 (https://doi.org/10.3168/jds.2020-18662). The slopes ranged from 0.30 for PY in Alpine when using the $WssGBLUP_{pseudo\text{-}SNP(DW)}$ method, to 1.21 for FU in Saanen when using the $ssGBLUP_{pseudo\text{-}SNP(LD)}$ method. The average (±SD) slopes for the Alpine breed, across all traits and scenarios, were as follows: 0.78 ± 0.16, 0.77 ± 0.17, 0.77 ± 0.18, 0.61 ± 0.16, 0.61 ± 0.15, and 0.58 ± 0.14, for $ssGBLUP_{pseudo\text{-}SNP(DW)}$, $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$, $ssGBLUP_{SNP}$, $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$, $WssGBLUP_{SNP}$, and $WssGBLUP_{pseudo\text{-}SNP(DW)}$, respectively. For the Saanen breed, the average (±SD) slopes, across all traits and scenarios, were as follows: 0.88 ± 0.18, 0.88 ± 0.19, 0.88 ± 0.20, 0.72 ± 0.15, 0.72 ± 0.15, and 0.70 ± 0.15, for $ssGBLUP_{pseudo\text{-}SNP(DW)}$, $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$, $ssGBLUP_{SNP}$, $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$, $WssGBLUP_{SNP}$, and $WssGBLUP_{pseudo\text{-}SNP(DW)}$, respectively. In general, fitting haplotypes as pseudo-SNP did not reduce the bias (inflation) of the GEBV estimates, but the differences were small when compared with $ssGBLUP_{SNP}$. On average, the WssGBLUP method resulted in more biased estimates compared with the other approaches.

DISCUSSION

The number of haplotypes defined across the genome with the DW method was the same between Alpine and Saanen goats. However, the number of pseudo-SNP was higher in the Alpine breed for haplotypes longer than 10 SNP. A greater genetic variability in Alpine, with 1.8% inbreeding against 2.8% inbreeding in Saanen (Carillier et al., 2014), could explain this result. No difference was detected in the number of pseudo-SNP for the LD_{hap} method. Carillier et al. (2013) have shown that LD is equal to 0.17 in Alpine and Saanen breeds for SNP spaced 50 kb apart. The LD in French dairy goats is smaller than the LD levels observed in dairy cattle (between 0.18 and 0.23; de Roos et al., 2008) but similar to other dairy goat and sheep populations (Brito et al., 2015, 2017). The LD between SNP in Alpine and Saanen were not high enough to create long haplotypes. In general, haplotypes were mainly shorter than 10 SNP. As a result, the number of alleles did not differ so much between Alpine and Saanen for this method. Karimi et al. (2018) also used pseudo-SNP for genomic predictions in dairy cattle. They used 21,236 Holstein phenotyped for 57 traits with a wide range of heritabilites (from 0.003 to 0.529). They classified the traits according to their heritability estimates into 3 classes: low (0–0.15), moderate (0.15–0.30), or high (>0.30). The haplotypes were constructed based on the DW method, including 5, 10, 15, and 20 SNP. In comparison with our study, they observed 75,263 pseudo-SNP for haplotypes of 5 SNP, which decreased to 37,270 pseudo-SNP for haplotypes containing 20 SNP, roughly half of the number of pseudo-SNP observed in French dairy goats. These results show that dairy goats present a higher genetic diversity than dairy cattle do, which is supported by the effective population size reported in the literature for both species (Brito et al., 2015). It is also important to mention that the SNP included in the SNP panel can affect the definition of haplotypes, as panels could have been designed after filtering out SNP in high LD or those that were not segregating in some breeds. This ascertainment bias could affect the haplotype definition and, consequently, the performance of genomic predictions.

The haplotypes constructed using the DW method led to a large number of pseudo-SNP. However, many were removed from the analyses because they were segregating at low frequency. In the LD_{hap} method, a large proportion of haplotypes was formed by individual SNP, because the LD estimates in French dairy goats were not high enough to combine SNP into haploblocks. An alternative would be to create fixed-length haplotypes, as described in dairy cattle by Hess et al. (2017). The authors evaluated fixed-length haplotypes (125 kb, 250 kb, 500 kb, 1 Mb, and 2 Mb) and reported improvement in GEBV accuracy for MY, FY, BW, and SCS with 250-kb haplotypes compared with genomic predictions based on individual SNP. The method proposed by Hess et al. (2017) could be useful to create haplotypes of different lengths, as in the LD_{hap} method, but limiting the number of haplotypes formed by individual SNP. If the size of the window is limited, it could also reduce the presence of many rare alleles, as was observed with the DW method.

In French dairy goats, previous genome-wide association studies using linkage and LD analysis (Martin et al., 2017, 2018) or $WssGBLUP_{SNP}$ (Teissier et al., 2018) identified QTL for milk yield traits, UFP, RUA, and SCS on CHI 19 in the Saanen breed. Other chromosomal regions of interest were located on CHI 6 for PC harboring the *α_{S1}-casein* gene (Grosclaude et al., 1987) and on CHI 14 for FC harboring the *DGAT1* gene (Martin et al., 2017) in both Alpine and Saanen breeds. Through the use of pseudo-SNP in $WssGBLUP_{pseudo\text{-}SNP(DW)}$ and $WssGBLUP_{pseudo\text{-}SNP(LD)}$, some of these regions were detected. Nevertheless, $WssGBLUP_{pseudo\text{-}SNP(DW)}$ and $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ did not detect the *DGAT1* gene for FY. This phenomenon was also observed in Teissier et al. (2018) with $WssGBLUP_{SNP}$. This could be due to the data set available. Indeed, animals genotyped with the 50K chip were almost all genotyped for the *DGAT1* mutations R251L and R396W (data not used in this study). The frequencies of the major alleles are between 76 and 99%, depending on the breed and mutation (re-

sults not shown). Almost no animals are homozygous for the minor alleles (less than 10 animals). Thus, WssGBLUP might not be able to correctly estimate SNP or pseudo-SNP effects associated with *DGAT1*. The use of pseudo-SNP allowed the detection of new chromosomal regions of interest, and, in particular, large pseudo-SNP weights were identified for FU in the Saanen breed on CHI 13 and CHI 19.

Some studies have reported improvement of accuracy with the use of haplotypes in genomic predictions (Calus et al., 2008; Cuyabano et al., 2014; Jónás et al., 2016). In these studies, haplotypes were not converted into pseudo-SNP. Karimi et al. (2018) also fitted haplotypes as pseudo-SNP. They compared $GBLUP_{SNP}$ and $GBLUP_{pseudo\text{-}SNP}$ with haplotypes constructed with the DW method for 5, 10, 15, and 20 SNP, and reported similar accuracies between $GBLUP_{SNP}$ and $GBLUP_{pseudo\text{-}SNP}$ for all the haplotype sizes and for the traits with low (0.20) and moderate heritability estimates (0.35). However, they observed only a slight improvement in GEBV accuracies for traits with moderate to high heritability ($h^2 > 0.20$) using $GBLUP_{pseudo\text{-}SNP}$ with 5 SNP (0.50) compared with $GBLUP_{SNP}$ (0.49). Accuracies of $GBLUP_{pseudo\text{-}SNP}$ with 10 (0.49), 15 (0.48), and 20 (0.47) SNP were lower than or statistically similar to those with $GBLUP_{SNP}$. In the current study, accuracies of $ssGBLUP_{pseudo\text{-}SNP}$ were breed- and trait-specific. For instance, higher gains in GEBV accuracy were observed for yield traits and small improvement for udder type traits in Alpine with the use of $ssGBLUP_{pseudo\text{-}SNP(DW)}$ or $ssGBLUP_{pseudo\text{-}SNP(LDhap)}$ compared with $ssGBLUP_{SNP}$. In Saanen, $ssGBLUP_{pseudo\text{-}SNP(DW)}$ was the best method for SCS.

This study is the first attempt to fit pseudo-SNP into the WssGBLUP method. In Saanen goats, $WssGBLUP_{pseudo\text{-}SNP(DW)}$ or $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ slightly outperformed or were as accurate as $WssGBLUP_{SNP}$ for milk yield and content traits, UFP, RUA, and SCS. These traits are known to have a QTL or major gene identified on various chromosomes (Martin et al., 2017, 2018). The $WssGBLUP_{pseudo\text{-}SNP}$ method was also able to outperform $ssGBLUP_{SNP}$, showing benefits of using pseudo-SNP in genomic evaluations. In Alpine goats, the use of $WssGBLUP_{pseudo\text{-}SNP(DW)}$ or $WssGBLUP_{pseudo\text{-}SNP(LDhap)}$ was the most interesting for FY, for which it outperformed both $ssGBLUP_{SNP}$ and $WssGBLUP_{SNP}$. Nevertheless, accuracies with pseudo-SNP were similar with and without the use of SNP weights. For the other traits, haplotype methods did not show advantage in terms of GEBV accuracy, especially compared with $ssGBLUP_{SNP}$.

The observed gains in accuracy in this study were higher than in other studies using haplotypes (Cuyabano et al., 2014; Karimi et al., 2018). This might be related to the genetic diversity of populations, especially effective population sizes and inbreeding. Inbreeding in dairy goats is generally lower (1.6 to 1.8%; Carillier et al., 2013) than in other livestock species, such as dairy cattle (>3%; Signer-Hasler et al., 2017; Doublet et al., 2019; Makanjuola et al., 2020). Lower inbreeding levels in goats also imply lower LD, and, therefore, haplotype-based methods might better capture the QTL effect and thus improve accuracy of GEBV.

A higher bias was observed with WssGBLUP compared with the other approaches. This may be expected due to the iterative nature of the WssGBLUP method, as variances are estimated from SNP effects from a previous iteration, leading to larger SNP effects in the next iteration and then to larger variances, and so on.

Genomic evaluations with haplotypes require several additional steps compared with genomic evaluations with individual SNP: (1) phasing the genotypes, (2) defining the haplotypes, (3) converting into pseudo-SNP, and (4) filtering according to their frequency. These steps make genomic evaluations with pseudo-SNP longer and more time-consuming than genomic evaluations based on individual SNP. With the LD method, it is also necessary to estimate LD between SNP. However, for some traits, it might be worth fitting haplotypes to obtain improvements in genomic predictive performance.

CONCLUSIONS

Genomic evaluations with pseudo-SNP improved the accuracy of genomic evaluations for some traits. With ssGBLUP using haplotypes fitted as pseudo-SNP, improvements up to +19% (0.310 to 0.368 for fat yield) and +3% (0.361 to 0.373 for udder shape) in Alpine and Saanen goats were observed, compared with ssGBLUP fitting individual SNP. The accuracy of genomic evaluations improved by a maximum of 24% (0.298 to 0.370 for fat yield in Alpine goats) and 14% (0.431 to 0.490 for SCS in Saanen) by using WssGBLUP with pseudo-SNP, compared with WssGBLUP with individual SNP. The average GEBV accuracy across both breeds and for ssGBLUP was equal to 0.50 with individual SNP and pseudo-SNP. For WssGBLUP, these average accuracies were equal to 0.50 with individual SNP and 0.49 with pseudo-SNP. This shows that the accuracies of these methods are trait dependent. We found that WssGBLUP was more biased than the ssGBLUP approach, and its gains in accuracy were limited to milk production traits. In general, fitting haplotypes as pseudo-SNP did not increase or reduce the bias of the GEBV estimates compared with single-SNP approaches. Despite the fact that genomic predictions based on haplotypes require additional steps and time, observed

gains in GEBV accuracy and bias reduction for some traits may be advantageous.

ACKNOWLEDGMENTS

This study would not have been possible without the goat SNP50 BeadChip developed by the International Goat Genome Consortium (IGGC): www.goatgenome.org. The authors also acknowledge Ignacy Misztal (University of Georgia, Athens, GA) for providing access to the blup90iod2 program, the French Genovicap and Phenofinlait programs [ANR (Paris, France), Apis-Gène (Paris), CASDAR, FranceAgriMer (Montreuil, France), France Génétique Elevage (Paris), the French Ministry of Agriculture Agrifood, and Forestry (Paris)], and the European 3SR Project, which funded part of this work. The first author also received financial support from the Occitanie region and the French National Institute for Agricultural Research (INRAE, Paris) SELGEN program (INCoMINGS). The authors have not stated any conflicts of interest.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730.

Brito, L. F., S. M. Clarke, J. C. McEwan, S. P. Miller, N. K. Pickering, W. E. Bain, K. G. Dodds, M. Sargolzaei, and F. S. Schenkel. 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 Genet. 18:7. https://doi.org/10.1186/s12863-017-0476-8.

Brito, L. F., M. Jafarikia, D. A. Grossi, J. W. Kijas, L. R. Porto-Neto, R. V. Ventura, M. Salgorzaei, and F. S. Schenkel. 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genet. 16:67. https://doi.org/10.1186/s12863-015-0220-1.

Broman, K. W., and J. L. Weber. 1999. Long homozygous chromosomal segments in reference families from the Centre d'Étude du Polymorphisme Humain. Am. J. Hum. Genet. 65:1493–1500. https://doi.org/10.1086/302661.

Calus, M. P., H. Huang, A. Vereijken, J. Visscher, J. ten Napel, and J. J. Windig. 2014. Genomic prediction based on data from three layer lines: A comparison between linear methods. Genet. Sel. Evol. 46:57. https://doi.org/10.1186/s12711-014-0057-5.

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561. https://doi.org/10.1534/genetics.107.080838.

Carillier, C., H. Larroque, I. Palhière, V. Clément, R. Rupp, and C. Robert-Granié. 2013. A first step toward genomic selection in the multi-breed French dairy goat population. J. Dairy Sci. 96:7294–7305. https://doi.org/10.3168/jds.2013-6789.

Carillier, C., H. Larroque, and C. Robert-Granié. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genet. Sel. Evol. 46:67. https://doi.org/10.1186/s12711-014-0067-3.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2.

Cuyabano, B. C., G. Su, and M. S. Lund. 2014. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. BMC Genomics 15:1171. https://doi.org/10.1186/1471-2164-15-1171.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503–1512. https://doi.org/10.1534/genetics.107.084301.

Doublet, A.-C., P. Croiseau, S. Fritz, A. Michenet, C. Hozé, C. Danchin-Burge, D. Laloë, and G. Restoux. 2019. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genet. Sel. Evol. 51:52. https://doi.org/10.1186/s12711-019-0495-1.

Edel, C., E. Pimentel, L. Plieschke, R. Emmerling, and K.-U. Götz. 2017. Effects of selective genotyping and selective imputation in single-step GBLUP. Interbull Bull. 51:22–25.

Feitosa, F. L. B., A. S. C. Pereira, S. T. Amorim, E. Peripolli, R. M. de Oliviera Silva, C. U. Braz, A. M. Ferrinho, F. S. Schenkel, L. F. Brito, R. Espigolan, L. G. de Albuquerque, and F. Baldi. 2020. Comparison between haplotype-based and individual snp-based genomic predictions for beef fatty acid profile in Nelore cattle. J. Anim. Breed. Genet. 137:468–476. https://doi.org/10.1111/jbg.12463.

Ferdosi, M. H., J. Henshall, and B. Tier. 2016. Study of the optimum haplotype length to build genomic relationship matrices. Genet. Sel. Evol. 48:75. https://doi.org/10.1186/s12711-016-0253-6.

Grosclaude, F., M.-F. Mahé, G. Brignon, L. Di Stasio, and R. Jeunet. 1987. A Mendelian polymorphism underlying quantitative variations of goat α_{S1}-casein. Genet. Sel. Evol. 19:399–412. https://doi.org/10.1186/1297-9686-19-4-399.

Hess, M., T. Druet, A. Hess, and D. Garrick. 2017. Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. Genet. Sel. Evol. 49:54. https://doi.org/10.1186/s12711-017-0329-y.

Hickey, J. M., B. P. Kinghorn, B. Tier, S. A. Clark, J. H. J. van der Werf, and G. Gorjanc. 2013. Genomic evaluations using similarity between haplotypes. J. Anim. Breed. Genet. 130:259–269. https://doi.org/10.1111/jbg.12020.

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299–1320. https://doi.org/10.1038/nature04226.

Jónás, D., V. Ducrocq, M.-N. Fouilloux, and P. Croiseau. 2016. Alternative haplotype construction methods for genomic evaluation. J. Dairy Sci. 99:4537–4546. https://doi.org/10.3168/jds.2015-10433.

Karimi, Z., M. Sargolzaei, J. A. B. Robinson, and F. S. Schenkel. 2018. Assessing haplotype-based models for genomic evaluation in Holstein cattle. Can. J. Anim. Sci. 98:750–759. https://doi.org/10.1139/cjas-2018-0009.

Larroque, H., J.-M. Astruc, A. Barbat, F. Barillet, D. Boichard, B. Bonaiti, I. David, G. Lagriffoul, I. Palhière, A. Piacère, C. Robert-Granié, and R. Rupp. 2011. National genetic evaluations in dairy sheep and goats in France. Page 62, Proc. Annual Meeting of the European Federation of Animal Science (EAAP). Aug. 29, 2011, Stavanger, Norway. Wageningen Academic Publishers, Wageningen, the Netherlands.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Makanjuola, B. O., F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J. Bertrand-Michel, I. Racke, H. Besir, R. Rupp, and G. Tosser-Klopp. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. Sci. Rep. 7:1872. https://doi.org/10.1038/s41598-017-02052-0.

Martin, P., I. Palhière, C. Maroteau, V. Clément, I. David, G. T. Klopp, and R. Rupp. 2018. Genome-wide association mapping for

type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. J. Dairy Sci. 101:5214–5226. https://doi.org/10.3168/jds.2017-13625.

Meuwissen, T. H., J. Odegard, I. Andersen-Ranberg, and E. Grindflek. 2014. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. Genet. Sel. Evol. 46:49. https://doi.org/10.1186/1297-9686-46-49.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. https://doi.org/10.3168/jds.2009-2064.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. G. Vitezica. 2016. Manual for BLUPF90 Family of Programs. University of Georgia, Athens, GA.

Piccoli, M. L., L. F. Brito, J. Braccini, H. R. Oliveira, F. F. Cardoso, V. M. Roso, M. Sargolzaei, and F. S. Schenkel. 2020. Comparison of genomic prediction methods for evaluation of adaptation and productive efficiency traits in Braford and Hereford cattle. Livest. Sci. 231:103864. https://doi.org/10.1016/j.livsci.2019.103864.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. https://doi.org/10.1086/519795.

Ricard, A., S. Danvy, and A. Legarra. 2013. Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. J. Anim. Sci. 91:1076–1085. https://doi.org/10.2527/jas.2012-5256.

Rogers, A. R., and C. Huff. 2009. Linkage disequilibrium between loci with unknown phase. Genetics 182:839–844. https://doi.org/10.1534/genetics.108.093153.

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. E. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6:39–44. https://doi.org/10.2527/af.2016-0006.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478.

Signer-Hasler, H., A. Burren, M. Neuditschko, M. Frischknecht, D. Garrick, C. Stricker, B. Gredler, B. Bapst, and C. Flury. 2017. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49:83. https://doi.org/10.1186/s12711-017-0358-6.

Teissier, M., H. Larroque, and C. Robert-Granié. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. Genet. Sel. Evol. 50:31. https://doi.org/10.1186/s12711-018-0400-3.

Teissier, M., H. Larroque, and C. Robert-Granie. 2019. Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. J. Dairy Sci. 102:3142–3154. https://doi.org/10.3168/jds.2018-15650.

Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadieu-Tonon, A. Eggen, H. C. M. Heuven, S. Jamli, A. J. Jiken, C. Klopp, C. T. Lawley, J. McEwan, P. Martin, C. R. Moreno, P. Mulsant, I. Nabihoudine, E. Pailhoux, I. Palhière, R. Rupp, J. Sarry, B. L. Sayre, A. Tircazes, J. Wang, W. Wang, W. Zhang, and the International Goat Genome Consortium. 2014. Design and characterization of a 52K SNP chip for goats. PLoS One 9:e86227. https://doi.org/10.1371/journal.pone.0086227.

Uemoto, Y., S. Sato, T. Kikuchi, S. Egawa, K. Kohira, H. Sakuma, S. Miyashita, S. Arata, T. Kojima, and K. Suzuki. 2017. Genomic evaluation using SNP- and haplotype-based genomic relationship matrices in a closed line of Duroc pigs. Anim. Sci. J. 88:1465–1474. https://doi.org/10.1111/asj.12805.

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

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. (Camb.) 94:73–83. https://doi.org/10.1017/S0016672312000274.

Zhang, X., D. Lourenco, I. Aguilar, A. Legarra, and I. Misztal. 2016. Weighting strategies for single-step genomic BLUP: An iterative approach for accurate calculation of GEBV and GWAS. Front. Genet. 7:151. https://doi.org/10.3389/fgene.2016.00151.

ORCIDS

Marc Teissier https://orcid.org/0000-0002-0137-961X
Luiz F. Brito https://orcid.org/0000-0002-5819-0922
Rachel Rupp https://orcid.org/0000-0003-3417-8336
Flavio S. Schenkel https://orcid.org/0000-0001-8700-0633
Christèle Robert-Granié https://orcid.org/0000-0001-5313-2187