

RESEARCH ARTICLE

Open Access



# 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

Marc Teissier\* , H el ene Larroque and Christ ele Robert-Grani e

## Abstract

**Background:** In 2017, genomic selection was implemented in French dairy goats using the single-step genomic best linear unbiased prediction (ssGBLUP) method, which assumes that all single nucleotide polymorphisms explain the same fraction of genetic variance. However, ssGBLUP is not suitable for protein content, which is controlled by a major gene, i.e.  $\alpha_{s1}$  *casein*. This gene explains about 40% of the genetic variation in protein content. In this study, we evaluated the accuracy of genomic prediction using different genomic methods to include the effect of the  $\alpha_{s1}$  *casein* gene.

**Methods:** Genomic evaluation for protein content was performed with data from the official genetic evaluation on 2955 animals genotyped with the Illumina goat SNP50 BeadChip, 7202 animals genotyped at the  $\alpha_{s1}$  *casein* gene and 6,767,490 phenotyped females. Pedigree-based BLUP was compared with regular unweighted ssGBLUP and with three weighted ssGBLUP methods (WssGBLUP, WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub>), which give weights to SNPs according to their effect on protein content. Two other methods were also used: trait-specific marker-derived relationship matrix (TABLUP) using pre-selected SNPs associated with protein content and gene content based on a multiple-trait genomic model that includes  $\alpha_{s1}$  *casein* genotypes. We estimated accuracies of predicted genomic estimated breeding values (GEBV) in two populations of goats (Alpine and Saanen).

**Results:** Accuracies of GEBV with ssGBLUP improved by +5 to +7 percent points over accuracies from the pedigree-based BLUP model. With the WssGBLUP methods, SNPs that are located close to the  $\alpha_{s1}$  *casein* gene had the biggest weights and contributed substantially to the capture of signals from quantitative trait loci. Improvement in accuracy of genomic predictions using the three weighted ssGBLUP methods delivered up to +6 percent points of accuracy over ssGBLUP. A similar accuracy was obtained for ssGBLUP and TABLUP considering the 20,000 most important SNPs. Incorporating information on the  $\alpha_{s1}$  *casein* genotypes based on the gene content method gave similar results as ssGBLUP.

**Conclusions:** The three weighted ssGBLUP methods were efficient for detecting SNPs associated with protein content and for a better prediction of genomic breeding values than ssGBLUP. They also combined fast computing, simplicity and required ssGBLUP to be run only twice.

\*Correspondence: marc.teissier@inra.fr  
GenPhySE, INRA, INPT, ENVT, Universit e de Toulouse,  
31326 Castanet-Tolosan, France



## Background

The availability of molecular data has enabled the development and commercial application of genomic selection in various livestock species, such as dairy cattle [1, 2], dairy sheep [3, 4], meat sheep [5, 6] and dairy goats [7–9]. Meuwissen et al. [10] proposed genomic prediction of animals based on dense single nucleotide polymorphism (SNP) maps, by deriving the effects of SNPs from a reference population, for which animals are both phenotyped and genotyped. Genomic estimated breeding values (GEBV) of selection candidates (i.e., usually young individuals with genotypes but without phenotypes) can be estimated by summing up the effects of the SNP alleles carried by each animal.

Methods such as genomic best linear unbiased prediction (GBLUP) [11–15], are used to predict GEBV by replacing the pedigree relationship matrix used for pedigree-based BLUP with a realized genomic relationship matrix. The GBLUP method was further improved with single-step GBLUP (ssGBLUP) [12], which uses simultaneously all phenotypic, pedigree and genotypic information, including phenotypic information on non-genotyped individuals. Therefore, in ssGBLUP, the relationship between each pair of animals (genotyped and non-genotyped) is estimated with a relationship matrix that combines pedigree and genotype information. Several studies have reported that the accuracy of genomic prediction obtained with these methods is higher than with genetic evaluation using pedigree-based BLUP [16–18]. However, the accuracy obtained from genomic information depends on several parameters including reference population size [19, 20], extent of linkage disequilibrium (LD), heritability of the trait [20, 21], relationship between training and validation populations [10] and the genetic architecture of the trait, which relates to the relative size of allele substitution effects at quantitative trait loci (QTL) [10, 22].

The GBLUP and ssGBLUP methods usually assume that each SNP follows the same distribution [11, 12, 16, 23–25], thus, all SNPs have the same variance and the same weight for SNP variance. However, different genomic evaluation methods have been developed to allow the variance of the effect of SNPs to differ between SNPs. A priori information can be used to modify the distribution of SNP effects. Giving more variance to some SNPs allows these methods to take the presence of major genes or QTL that affect the trait of interest into account. For instance, various Bayesian methods, which estimate the effect of SNPs from animals that are both genotyped and phenotyped, have been proposed [10, 26–28]. The main difference between these Bayesian methods lies in the definition of an a priori distribution of the effects of SNPs. SNPs can be attributed to different

distribution classes, which explain different parts of the total genetic variance, with one class possibly containing the SNPs that have no effect on the trait. Because animals need to be phenotyped and genotyped to apply Bayesian methods, phenotypes from non-genotyped animals cannot be included. In dairy breeding programs, genotypes are mainly determined on the males whereas phenotypes come from the females. Thus, daughter yield deviations (DYD) or de-regressed proofs are calculated to obtain pseudo-phenotypes for the males. However, multi-step methods may create bias in genomic predictions [29].

Other methods based on the ssGBLUP framework such as weighted ssGBLUP (WssGBLUP) or on the trait-specific marker-derived relationship matrix (TABLUP) have been proposed [30]. WssGBLUP is an extension of ssGBLUP in which weights for SNP variances are used when forming the genomic relationship matrix [12]. WssGBLUP can set more weight to SNPs that are in high LD with a causal mutation or associated with QTL with a relatively large effect. These weights are estimated from the variance explained by each SNP as presented by Wang et al. [23]. The weighting of SNP variances was also investigated by Zhang et al. [24] who proposed to use the same weight for SNPs that are within a defined window along the genome. The TABLUP method proposes to construct the genomic relationship matrix based on genotypes from a subset of pre-selected SNPs. Selection of SNPs can be performed after GWAS analysis or based on weights that are estimated with WssGBLUP. The selected SNPs are then equally weighted for the analyses [30]. Furthermore, an alternative to the previous methods is the gene content method proposed by Gengler et al. [31], which is based on a multiple trait model and considers the gene content for specific genotypes as a new trait. This method can combine information from SNPs and genotypes for a causal mutation [31, 32]. The number of alleles carried by each animal is considered as a second trait correlated to the quantitative trait. Then, the causal mutation is integrated directly in the ssGBLUP multiple-trait model. Its advantage is that it can be extended to multi-allelic genes and used when genotypes for a causal mutation are missing [33].

In French dairy goats, the first step towards genomic selection for milk production traits, udder type traits and somatic cell score was taken by Carillier-Jacquin et al. [8, 9] for French Alpine and Saanen dairy goat breeds. Carillier-Jacquin et al. [8, 9] compared ssGBLUP and other methods of genomic evaluation that require several steps (GBLUP or Bayesian methods). GBLUP and Bayesian methods usually use performances based on pseudo-phenotypes (DYD) whereas ssGBLUP is based on female performance. These authors found that ssGBLUP gave more accurate predictions of the genetic merit

of selection candidates than the previous official genetic evaluation that did not use genomic information, or the use of multi-step genomic methods. However, the increase in accuracy due to using genomic information was not expected to be high because the reference population was small.

Currently, the next step in the genomic evaluation of French dairy goats is to investigate better ways to use genotyping information to improve the accuracy of genomic evaluation. One possibility is to take prior knowledge about major genes into account. Several major genes have been identified, such as *DGATI* for fat content [34] and  $\alpha_{s1}$  *casein* for protein content [35]. For protein content, Carillier-Jacquin et al. [33] reported that the genetic variance explained by the  $\alpha_{s1}$  *casein* gene reached 38% in the Saanen and 43% in the Alpine breed. The caprine  $\alpha_{s1}$  *casein* gene has six alleles (*A*, *B*, *C*, *E*, *F* and *O*) that have been identified in the French dairy goat population. Allele *A* is predominant in the Alpine breed, whereas alleles *A*, *E* and *F* are the most frequent in the Saanen breed [33]. Carillier-Jacquin et al. [33] showed that integrating the  $\alpha_{s1}$  *casein* gene for protein content with the gene content method improved the accuracy of genomic evaluation (+8 to 14% for Alpine and Saanen populations) compared with ssGBLUP.

In this study, our aim was to investigate different methods of genomic prediction that estimate and integrate the fact that chromosomal regions are strongly associated with a trait. Protein content in French dairy goats was analyzed by applying WssGBLUP, two alternatives of the WssGBLUP method, the TABLUP method and the gene content method. These methods were compared with pedigree-based BLUP and ssGBLUP based on the accuracies of predicted breeding values.

## Methods

### Animals, phenotypes and genotypes

The dataset used in this study was provided by the French national milk records system and included animals from the two main French dairy goat breeds, Alpine and Saanen. Phenotypes for protein content, pedigree data, genotypes and environmental fixed effects used in the ssGBLUP method were obtained from the official genetic evaluation of January 2016 [36]. Analyses were performed

with a multi-breed dataset (Alpine and Saanen animals combined) and in two separate within-breed analyses.

The trait analyzed was protein content (g/kg) with measurements from 6,767,490 lactations and 2,458,453 females recorded between 1980 and 2010. Descriptive statistics (animal and record numbers, minimum, mean, maximum, coefficient of variation) for each breed are in Table 1.

The pedigree consisted of 2,543,789 animals (1,449,991 Alpine and 1,093,798 Saanen). In addition, it was completed with 36 unknown parent groups. Unknown parent groups were defined for each breed and for animals born before 1975, and then for cohorts born in 2-year windows up to 2010.

Animals that were genotyped with the Illumina goat SNP50 BeadChip (50K SNP) [37] were also used in the analysis. Quality control (QC) for a dataset of 3347 genotyped animals (2020 Alpine and 1278 Saanen) and 53,347 SNPs was performed independently for each breed. SNPs with a minor allele frequency (MAF) lower than 1% and a call rate lower than 95% were removed. Hardy-Weinberg equilibrium was also tested and the associated Chi squared statistic was calculated for each SNP. SNPs with a Chi squared statistic higher than 24 were removed. Finally, animals with a SNP call rate lower than 99% were discarded from the analyses. After QC, 2955 (1749 Alpine and 1206 Saanen) animals and 46,849 SNPs remained for further analyses. Some SNPs within the  $\alpha_{s1}$  *casein* gene were present on the 50 K SNP but since they did not pass QC, they were removed [33].

Genotypes for the  $\alpha_{s1}$  *casein* gene were available for 3696 Alpine individuals (2154 males and 1542 females), and 3506 Saanen individuals (2049 males and 1457 females) born between 1982 and 2012. The  $\alpha_{s1}$  *casein* gene is located on caprine chromosome 6 at 82 Mb and is multi-allelic in the French dairy goat population, with six different alleles (*A*, *B*, *C*, *E*, *F* and *O*) and 19 genotypes detected among the 21 possibilities (*FO* and *OO* genotypes have never been detected in the French dairy goat population) [33]. Genotypes of animals with one missing allele were removed from the analysis. The estimated effects of the 19  $\alpha_{s1}$  *casein* genotypes on protein content were computed and reported previously [33]. Table 2 includes the number of animals (males and females for

**Table 1 Summary statistics on protein content (g/kg) in Alpine and Saanen breeds**

| Breed  | Number of lactations | Number of females with phenotypes | Minimum <sup>a</sup> (g/kg) | Mean <sup>a</sup> (g/kg) | Maximum <sup>a</sup> (g/kg) | CV   |
|--------|----------------------|-----------------------------------|-----------------------------|--------------------------|-----------------------------|------|
| Alpine | 3,844,071            | 1,392,399                         | 10.47                       | 30.42                    | 54.81                       | 0.11 |
| Saanen | 2,923,419            | 1,066,054                         | 10.00                       | 29.67                    | 54.63                       | 0.09 |

CV coefficient of variation

<sup>a</sup> Minimum, mean, maximum protein content

**Table 2** Number of animals with information on the  $\alpha_{s1}$  casein genotype and/or 50 K SNP genotypes

| Breed  | Gender  | Animals with 50 K SNP genotype | Animals with $\alpha_{s1}$ casein genotype | Animals with both 50K SNP and $\alpha_{s1}$ casein genotype |
|--------|---------|--------------------------------|--|---|
| Alpine | Males   | 512                            | 2154                                       | 510   |
|        | Females | 1237                           | 1542                                       | 0   |
| Saanen | Males   | 393                            | 2049                                       | 393   |
|        | Females | 813                            | 1457                                       | 0   |

Alpine and Saanen breeds) used in this study with information on their  $\alpha_{s1}$  casein and/or 50 K SNP genotypes.

### Genomic prediction with and without considering information on the $\alpha_{s1}$ casein genotypes

ssGBLUP was implemented in 2017 in the official genetic evaluations for the two main French dairy goats. This method and pedigree-based BLUP were used as the reference method in our study and compared with WssGBLUP, two alternatives of the WssGBLUP method, TABLUP and the gene content method. Analyses were performed using the blupf90 software [38].

### Single-step GBLUP (ssGBLUP) method

For both multi-breed and within-breed scenarios, the following model was applied:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e}, \quad (1)$$

where  $\mathbf{y}$  is a vector of performances (female phenotypes) for protein content (phenotypes are based on standardized 250-day lactation records).  $\boldsymbol{\beta}$  is a vector of fixed effects including herd within year (32 years from 1980 to 2012) and within parity (1, 2 and  $\geq 3$ ) (188,933 levels in total); age at delivery within year and within region (four regions in France depending on goat breeding management) (3224 levels in total); month at delivery within year and region (1448 levels in total); and length of dry period within year and region (1107 levels in total); a fifth fixed effect for breed (two levels) was added for multi-breed analyses.  $\mathbf{u}$  is a vector of random additive genetic effects assumed to be normally distributed  $N(0, \mathbf{H}\sigma_u^2)$ ,  $\mathbf{p}$  is a vector of random permanent environmental effects assumed to be normally distributed  $N(0, \mathbf{I}\sigma_p^2)$ ,  $\mathbf{e}$  is a vector

of random residuals that is normally distributed  $N(0, \mathbf{I}\sigma_e^2)$ .  $\mathbf{X}$  is the incidence matrix relating phenotypes to the fixed effects ( $\boldsymbol{\beta}$ ).  $\mathbf{Z}$  is the design matrix allocating phenotypes to breeding values ( $\mathbf{u}$ ) and  $\mathbf{W}$  is the incidence matrix relating phenotypes to permanent environmental effects ( $\mathbf{p}$ ).

Matrix  $\mathbf{H}$  is the genetic relationship matrix combining SNP information and pedigree data, implemented as in Legarra et al. [12]:

$$\mathbf{H} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{pmatrix},$$

where  $\mathbf{A}$  is a pedigree-based relationship matrix with indices 1 for ungenotyped animals and 2 for genotyped animals, and  $\mathbf{G}$  is the genomic relationship matrix derived as in Christensen and Lund [11]:

$$\mathbf{G} = 0.95 \frac{\mathbf{M}'\mathbf{M}}{2 \sum_{i=1}^m p_i(1-p_i)} + 0.05\mathbf{A}_{22},$$

where  $m$  is the number of SNPs,  $p_i$  is the estimated allele frequency at locus  $i$  and  $\mathbf{M}$  is a centered matrix of SNP genotypes.

Variance components were estimated by using the restricted maximum likelihood (REML) method in the remlf90 software [38].

### Weighted ssGBLUP (WssGBLUP) method

Model 1 was also used for WssGBLUP but  $\mathbf{G}$  was constructed differently. Solutions of genomic breeding values from ssGBLUP (Model 1) can be decomposed into SNP effects as modeled in Wang et al. [23]:

$$\hat{\mathbf{a}} = \mathbf{D}\mathbf{M}'[\mathbf{MDM}']^{-1}\hat{\mathbf{u}}_g,$$

where  $\hat{\mathbf{a}}$  is a vector of SNP effects,  $\mathbf{D}$  is a diagonal matrix of weights (initially diagonal of 1 for the ssGBLUP),  $\mathbf{M}$  is the centered matrix of SNP genotypes and  $\hat{\mathbf{u}}_g$  the vector of GEBV from genotyped animals only. Variances of the effect of SNP  $i$  were estimated as:

$$\sigma_{u,i}^2 = 2\hat{a}_i^2 p_i(1-p_i),$$

where  $p_i$  is the allele frequency of SNP  $i$ . The vector of variances of SNP effects was normalized (the normalization process ensured that the sum of the variances remained constant and was equal to the number of SNPs) and used as weights in matrix  $\mathbf{D}$  to construct the weighted matrix  $\mathbf{G}$  ( $\mathbf{G}^*$ ) as described in Wang et al. [23]:



$$\mathbf{G}^* = 0.95 \frac{\mathbf{M}'\mathbf{D}\mathbf{M}}{2 \sum_{i=1}^m p_i(1-p_i)} + 0.05\mathbf{A}_{22}.$$

GEBV were estimated again with Model 1 by considering weights for each SNP via the  $\mathbf{G}^*$  matrix included in the  $\mathbf{H}$  matrix. This process was carried out iteratively with weights estimated at each iteration as described in Wang et al. [23]. Wang et al. [23] have shown that WssGLUP with only very few iterations may be sufficient to reach a maximum accuracy of GEBV and SNP effects. In this study, we analyzed the influence of the number of iterations (1–10) on the accuracy of genomic predictions.

As proposed by Zhang et al. [24], other methods can be considered to calculate the weight for SNPs in the  $\mathbf{D}$  matrix. These methods assign the same weight to several consecutive SNPs within a chromosomal region. Modifications of the WssGBLUP method were considered in this study and the individual weights were computed as follows: (1) the maximum weight of SNPs included in the chromosomal region, or (2) the sum of the weights of the SNPs included in the chromosomal region. These weights were calculated based on the weights estimated with the WssGBLUP. In the end, the vector of the weights was normalized in such a way that the sum of all weights remained constant and equal to the number of SNPs. Chromosomal regions of various lengths were tested: 2, 5, 10, 20, 40, 80, 100, 150, 200 and 250 consecutive SNPs with non-overlapping windows. Hereafter, these methods are named WssGBLUP<sub>*i*</sub> where *i* denotes the method used to calculate the weights (Max or Sum).

#### Trait-specific marker-derived relationship matrix (TABLUP) method

Only a subset of SNPs that are more or less associated with protein content was selected to build the  $\mathbf{G}$  matrix. One of our objectives was to investigate how the genetic architecture of protein content could be taken into account in the ssGBLUP method. Thus, TABLUP was applied by selecting a subset of SNPs according to their effect on the trait (estimated from the WssGBLUP method described previously). A total of 5000, 10,000, 15,000, 20,000, 25,000, 30,000, 35,000 or 40,000 SNPs were selected to construct  $\mathbf{G}$ . Two scenarios were tested in which either the most or the least strongly associated SNPs were selected. GEBV were estimated with Model 1 and the  $\mathbf{G}$  matrix that was built based on the selected SNPs without weights ( $\mathbf{D} = \mathbf{I}$ ).

#### Gene content method

The gene content method estimates the GEBV for each animal by taking information on the  $\alpha_{sI}$  casein genotype, genotypes from the 50K SNP and pedigree into account

through a multiple-trait model. The model used here was the same as in [33]:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e} \\ \mathbf{y}_A &= \mu_A + \mathbf{Z}_A\mathbf{u}_A + \mathbf{e}_A \\ \mathbf{y}_B &= \mu_B + \mathbf{Z}_B\mathbf{u}_B + \mathbf{e}_B \\ \mathbf{y}_C &= \mu_C + \mathbf{Z}_C\mathbf{u}_C + \mathbf{e}_C, \\ \mathbf{y}_E &= \mu_E + \mathbf{Z}_E\mathbf{u}_E + \mathbf{e}_E \\ \mathbf{y}_F &= \mu_F + \mathbf{Z}_F\mathbf{u}_F + \mathbf{e}_F \\ \mathbf{y}_O &= \mu_O + \mathbf{Z}_O\mathbf{u}_O + \mathbf{e}_O \end{aligned} \quad (2)$$

where  $\mathbf{y}$  is a vector of female performances for protein content. Fixed effects ( $\boldsymbol{\beta}$ ), random effects ( $\mathbf{u}$ ,  $\mathbf{p}$  and  $\mathbf{e}$ ) and incidence matrices  $\mathbf{X}$ ,  $\mathbf{Z}$  and  $\mathbf{W}$  are the same as in Model 1.  $\mathbf{y}_A$ ,  $\mathbf{y}_B$ ,  $\mathbf{y}_C$ ,  $\mathbf{y}_E$ ,  $\mathbf{y}_F$ , and  $\mathbf{y}_O$  are vectors of gene content for alleles *A*, *B*, *C*, *E*, *F* and *O*. This corresponds to the number of copies carried by each animal (i.e., 0, 1 or 2). For ungenotyped animals, the value was set to missing.  $\mu_A$ ,  $\mu_B$ ,  $\mu_C$ ,  $\mu_E$ ,  $\mu_F$ , and  $\mu_O$  are the mean fixed effects for alleles *A*, *B*, *C*, *E*, *F* and *O*,  $\mathbf{Z}_A$ ,  $\mathbf{Z}_B$ ,  $\mathbf{Z}_C$ ,  $\mathbf{Z}_E$ ,  $\mathbf{Z}_F$ , and  $\mathbf{Z}_O$  are the incidence matrices relating observations to the random genetic effect ( $\mathbf{u}_A$ ,  $\mathbf{u}_B$ ,  $\mathbf{u}_C$ ,  $\mathbf{u}_E$ ,  $\mathbf{u}_F$  and  $\mathbf{u}_O$ ) of gene content for each allele and  $\mathbf{e}_A$ ,  $\mathbf{e}_B$ ,  $\mathbf{e}_C$ ,  $\mathbf{e}_E$  and  $\mathbf{e}_O$  are the random residual errors for each of the six alleles. For  $i \in \{A, B, C, E, F, O\}$ ,  $\mathbf{u}_i$  are normally distributed such that  $\text{Var}(\mathbf{u}_i) = \mathbf{H}\sigma_{u_i}^2$  and  $\sigma_{u_i}^2 = 2p_i(1-p_i)$ , where  $p_i$  is the frequency of allele *i* at the  $\alpha_{sI}$  casein locus. Covariances between genetic values ( $\mathbf{u}$ ) and genetic effects of gene content ( $\mathbf{u}_A$ ,  $\mathbf{u}_B$ ,  $\mathbf{u}_C$ ,  $\mathbf{u}_E$ ,  $\mathbf{u}_F$  and  $\mathbf{u}_O$ ) were modeled as in Carillier-Jacquin et al. [33]. Variance and covariance parameters from this model were estimated using the restricted maximum likelihood (REML) algorithm implemented in the remlf90 software.

#### Accuracy of genomic predictions

Genomic evaluations were performed from all phenotypes recorded until January 2010, but we were also interested in the prediction of genotyped animals that constituted our reference population. This reference population was composed of 905 sires born between 1993 and 2012 and genotyped with the 50K SNP chip (Table 2) and was split into a training population of 554 sires born from 1993 to 2007 (307 Alpine and 247 Saanen) with phenotypes of their daughters recorded until January 2010), and 351 validation sires born from 2008 to 2012 (205 Alpine and 146 Saanen) with no daughters in January 2013 (daughters of these animals were removed from the dataset). Then, GEBV and DYD computed from the official genetic evaluation of January 2016 were compared for the 351 animals in the validation set. DYD were average performance values for the daughters corrected

for environmental effects and merit of the dam, and they were weighted by effective daughter contributions as described in VanRaden and Wiggans [39]. Accuracy of genomic predictions was assessed as the Pearson correlation between GEBV estimated with each model and DYD. Pearson correlations obtained with different methods were tested using the Hotelling-Williams test [40].

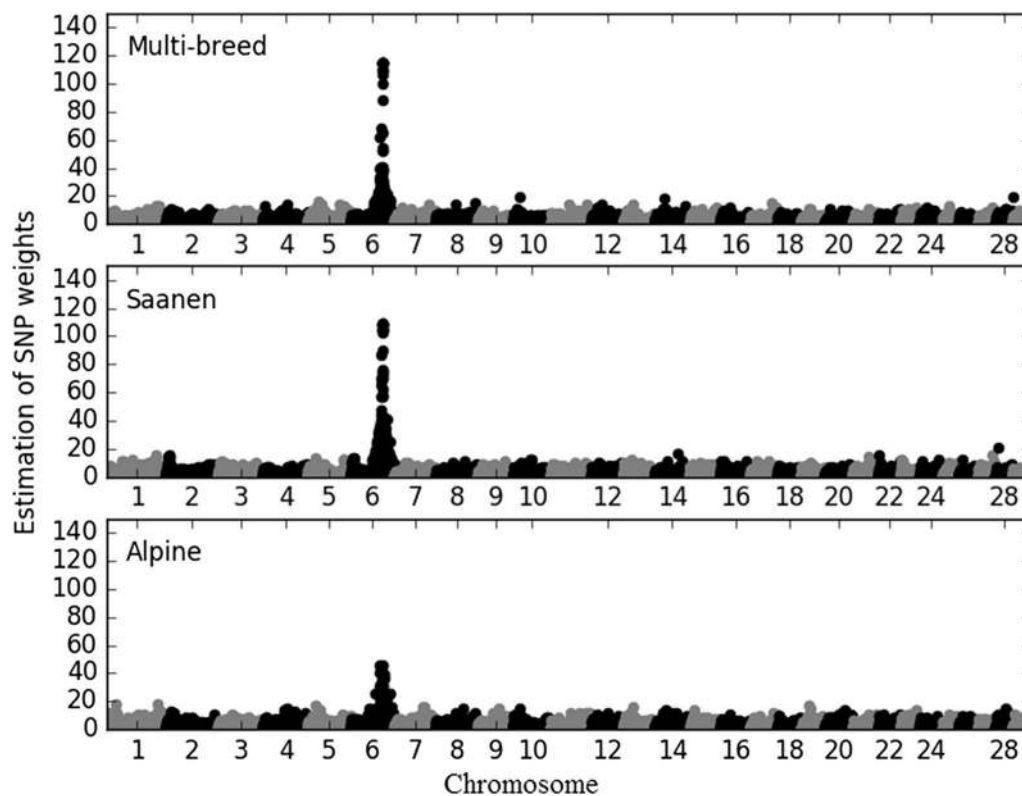
## Results and discussion

The most frequent  $\alpha_{s1}$  casein genotypes are *AA* for the males and *AE* for the females in the Alpine breed, and *AE* for the females and *EE* for the males in the Saanen breed (present in more than 50% of the animals). Allele *C* is rather rare (less than 5% of the animals carry this allele) in the two breeds. The largest differences in genotype frequency between Alpine and Saanen populations were observed for genotypes *AA* (49% in Alpine vs. 7% in Saanen), *EE* (3% in Alpine vs. 32% in Saanen) and *AE* (49% in Saanen vs. 30% in Alpine). These results were consistent with the previous work of Carillier-Jacquin et al. [33] in which fewer genotypes were available. Protein content was analyzed knowing that this trait is highly

heritable in both Alpine and Saanen populations (0.5) [41].

### Estimation of weights for SNPs with the WssGBLUP method

We compared different genomic methods. First, we used WssGBLUP because we wanted to identify the weights given to SNPs with this method, in order to determine if the chromosomal region including the  $\alpha_{s1}$  casein gene was considered in the analyses. WssGBLUP is an iterative method, and 10 iterations were performed for multi-breed analyses and within-breed analyses. Accuracy of genomic predictions was evaluated at each iteration (results not shown). The highest accuracies were obtained at the second iteration as reported by Wang et al. [23] and then decreased slightly. Thus, all the results presented for the WssGBLUP multi-breed and within-breed analyses are those obtained for the second iteration (see Fig. 1). The top 50 SNPs (with the biggest weights) were compared between the three analyses and were all located on chromosome 6 i.e. the multi-breed (between 71 and 86 Mb), Alpine (between 64 and 101 Mb) and Saanen analyses (between 71 and 92 Mb), and their weights ranged from 24 to 115 for multi-breed, from 23 to 45 for



**Fig. 1** Estimated weights of SNPs included in the second iteration of the WssGBLUP approach for multi-breed, Alpine and Saanen populations. WssGBLUP used GEBV of genotyped animals and genotypes to estimate weights for each SNP. The estimated weight of each SNP is represented along the genome

Alpine and from 30 to 108 for Saanen analyses. Among these SNPs, 16 were common to the three analyses and located between 78 and 82 Mb; 11 SNPs were common to the Saanen and multi-breed analyses and located between 79 and 83 Mb; 16 SNPs were common to the Alpine and multi-breed analyses and located between 77 and 86 Mb; and only one SNP was common to both the Alpine and Saanen analyses and located at 76 Mb.

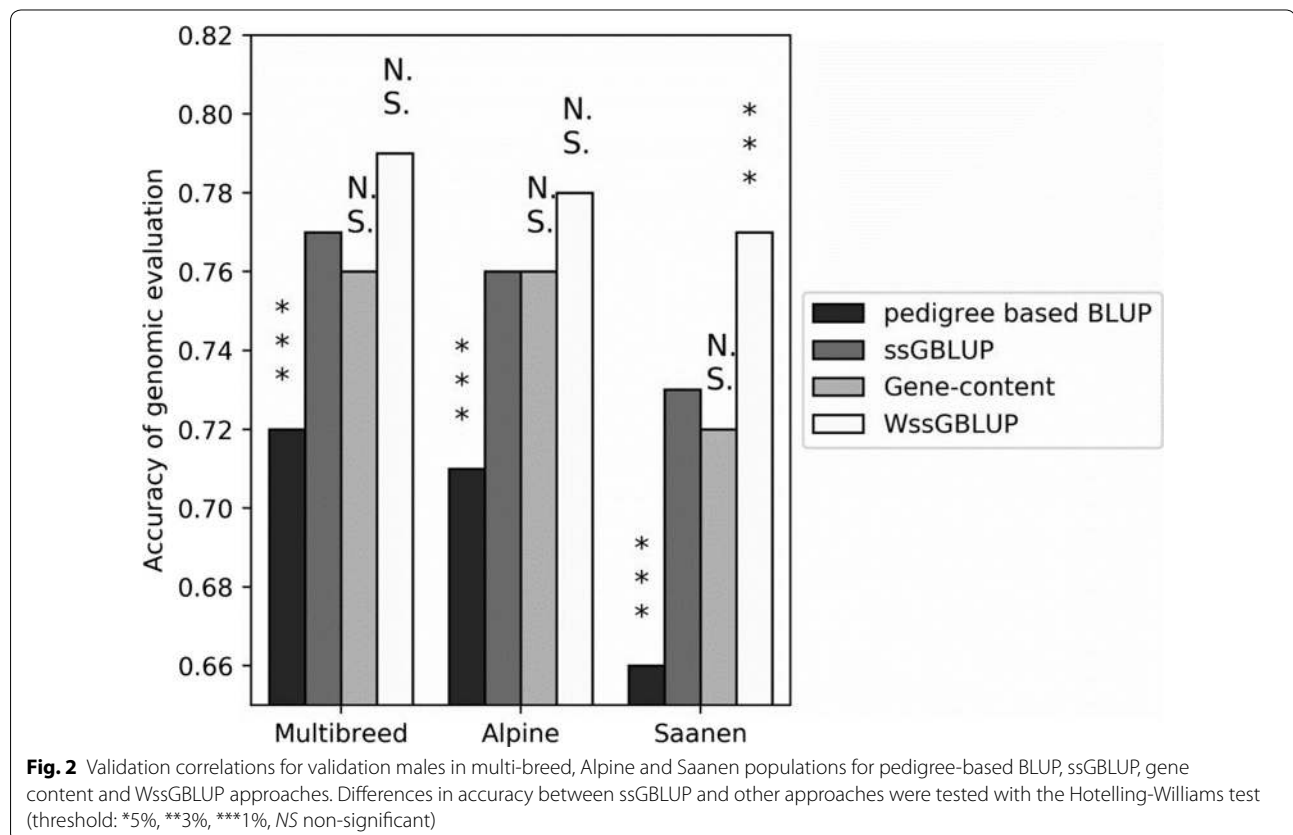
WssGBLUP can be used not only for genomic prediction but also for QTL detection as in GWAS [23, 24]. In French dairy goat data, the chromosomal regions detected with WssGBLUP were on caprine chromosome 6, which includes a well-known region that was previously located and described by Martin et al. [34] in a GWAS study. They performed linkage analyses (LA) and linkage disequilibrium (LD) analyses on 1941 dairy goats distributed in 20 half-sib families using all females and their 20 sire genotypes and detected a large QTL between 82.5 and 82.8 Mb on chromosome 6. In our study, SNPs with the biggest weights for SNP variances were located within this region.

The WssGBLUP method developed by Wang et al. [23] has some limitations. Weights for SNP variances are estimated by using a whole-genome regression, which can result in their unstable prediction due to

multi-collinearity between SNPs because of LD between SNPs. In our study, we tested common weights for several SNPs instead of individual weights for SNP variances, using WssGBLUP<sub>Max</sub> or WssGBLUP<sub>Sum</sub>. These methods are expected to limit the large variation in prediction of weights for SNP variances by smoothing weights of SNPs that are in the same window. In our study, WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> gave higher accuracies of genomic prediction than the classical WssGBLUP. With WssGBLUP<sub>Max</sub> or WssGBLUP<sub>Sum</sub>, window sizes were used to allocate the same weights to consecutive SNPs. Another approach would be to use the LD between SNPs, which could limit the multi-collinearity between the SNPs used in the genomic evaluation. Since the weight of SNPs is included through the **D** matrix, this matrix can be replaced by the weights derived from the GWAS approach.

#### Including the effect of the $\alpha_{s1}$ casein gene in WssGBLUP or gene content methods

Figure 2 presents accuracies of genomic evaluation for pedigree-based BLUP, ssGBLUP, gene content and WssGBLUP in a multi-breed population and in the Alpine and Saanen breeds. Accuracies with pedigree-based BLUP (0.72 in multi-breed, 0.71 in Alpine and 0.66



in Saanen) were lower than accuracies with ssGBLUP (0.77 in multi-breed, 0.76 in Alpine and 0.73 in Saanen), gene content (0.76 in multi-breed, 0.76 in Alpine and 0.72 in Saanen) or WssGBLUP (0.79 for multi-breed, 0.78 for Alpine and 0.77 for Saanen). The gene content method did not improve accuracy of genomic predictions for the three populations compared to ssGBLUP (accuracy was 1 percent point lower for gene content in the multi-breed and Saanen analyses and identical in the Alpine analysis). In addition, accuracies with WssGBLUP were significantly higher than with ssGBLUP for the Saanen population (+4 percent points). We did not observe any significant difference between ssGBLUP and WssGBLUP for multi-breed and Alpine populations.

Previously, Carillier-Jacquin et al. [33] used the gene content and ssGBLUP methods to analyze protein content in French dairy goats. Accuracies obtained with ssGBLUP were higher in our study than in Carillier-Jacquin et al. [33] for the multi-breed (+5 percent points) and Alpine (+8 percent points) analyses, and slightly lower for the Saanen analysis (−2 percent points). A similar trend was observed with the gene content method, with +1 percent point for multi-breed, +8 percent points for Alpine and −14 percent points for Saanen in our study compared to Carillier-Jacquin et al. [33]. The main difference between our study and that of Carillier-Jacquin et al. [33] was the number of animals genotyped with the 50 K SNP chip, number of  $\alpha_{s1}$  *casein* genotypes, and the size and composition of the training and validation sets. In our study, 82 males and 2050 females genotyped with the 50 K SNP chip and 50 females and 878 males genotyped for the  $\alpha_{s1}$  *casein* gene were added. In Carillier-Jacquin et al. [33], the reference population consisted of a training set with 677 animals born between 1993 and 2009 (384 Alpine and 293 Saanen), and a validation set with 146 animals born between 2010 and 2011 (86 Alpine and 60 Saanen). In our study, we had 554 animals born between 1993 and 2007 (307 Alpine and 247 Saanen) in the training set and 351 animals born between 2008 and 2012 (205 Alpine and 146 Saanen) in the validation set. The main difference between the Carillier-Jacquin et al. study and that reported here was the size of the validation population (2 versus 5 years in our study). The slightly improved results that we obtained may be explained by the larger reference population (823 animals in Carillier-Jacquin et al. [33] compared to 905 in our study), a well-known factor in the literature on genomic selection. For instance, VanRaden et al. [42] report a gain of +5 percent points between genomic prediction and parent average by adding 1000 animals in the training population. These results were consistent with the higher accuracy obtained in the multi-breed analysis compared to the within-breed

analyses, especially if the trait has the same genetic determinism in the two breeds that are combined (which is the case for protein content). Accuracy is expected to improve even more the size of the reference population continues to grow over the years.

Carillier-Jacquin et al. [33] showed that the gene content method was more accurate than ssGBLUP (+3 percent points for multi-breed, +5 percent points for Alpine and +11 percent points for Saanen). However, in our study, accuracies of genomic prediction were the same for the gene content method and ssGBLUP. The goat  $\alpha_{s1}$  *casein* gene has six alleles in the two main French dairy goats and genotype frequencies vary considerably with some being rare. Predicting  $\alpha_{s1}$  *casein* genotypes with the gene content method for non-genotyped animals remains difficult in this case, especially in French dairy goats, for which the number of non-genotyped animals is large compared with that of genotyped animals (only 0.3% of the population is genotyped for the  $\alpha_{s1}$  *casein* gene), and 40% of females have unknown parents. This may explain why the gene content method did not outperform ssGBLUP.

The genetic architecture of protein content is similar between the Alpine and Saanen breeds. However, the gain in accuracy with the genomic evaluation methods (ssGBLUP, gene content and WssGBLUP) compared to pedigree-based BLUP was greater for the Saanen than the Alpine breed. As discussed by Carillier-Jacquin et al. [9], the greater gain observed for the Saanen breed between pedigree-based BLUP and genomic evaluation may be explained by a higher level of inbreeding (2.3% in Saanen and 1.8% in Alpine), and a higher kinship coefficient between the training and validation sets (2.4% in Saanen and 1.1% in Alpine using genomic data).

For prediction of GEBV, WssGBLUP was more efficient than gene content, which may be due to the construction of the 50K SNP chip. The region around the  $\alpha_{s1}$  *casein* gene was enriched in SNPs in the 1-Mb region at 82 Mb on chromosome 6 (the region that contains the  $\alpha_{s1}$  *casein* gene). Overall, 40 SNPs are present within this 1-Mb region, whereas on average only 20 SNPs per Mb are located outside of this region on chromosome 6 or on other chromosomes. Moreover, the Chi squared test between  $\alpha_{s1}$  *casein* genotypes and each SNP on chromosome 6 revealed a very strong correlation between  $\alpha_{s1}$  *casein* genotypes and SNPs on the 50K SNP chip in this region (results not shown). Giving more weight to SNPs that are more strongly associated with protein content seems to be more efficient to capture the effect of the  $\alpha_{s1}$  *casein* gene than using genotype data for this gene. Vallejo et al. [18] investigated the efficiency of WssGBLUP for bacterial cold water disease resistance, for which several QTL are identified. They observed an improvement of 4



percent points with WssGBLUP compared to ssGBLUP. In our study, we observed similar gains with WssGBLUP. Su et al. [43] also observed a superiority of the WssGBLUP over ssGBLUP in dairy cattle for milk traits.

#### Use of common weights on consecutive SNPs with WssGBLUP

WssGBLUP was significantly more predictive than other genomic evaluation methods for protein content in the Saanen breed but not in multi-breed or the Alpine breed. Zhang et al. [24] reported that WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> increase the accuracy of genomic evaluation more efficiently than WssGBLUP. We evaluated these methods and Tables 3, 4 and 5 show the results on the validation population in the multi-breed, Alpine and Saanen populations, respectively using WssGBLUP and the two modified WssGBLUP methods (Max, Sum) according to the size of SNP windows. If identical results were obtained for different window sizes, they were merged in the same column. For the multi-breed population, accuracies of the analyses with WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> were very similar and differed only with non-overlapping SNP windows of 40, 80, 100, 150, 200 and 250 SNPs, the accuracy (0.81) of WssGBLUP<sub>Sum</sub> being slightly higher than that of WssGBLUP<sub>Max</sub> (0.80). Otherwise, accuracies were equal to 0.79 with a window size of two SNPs and 0.80 for window sizes of five, 10 and 20 SNPs. Finally, accuracies of WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> were slightly higher than that of WssGBLUP (0.79) and higher than that of ssGBLUP (0.77).

For both within-breed analyses, increasing the window size barely influenced accuracies. In the Alpine within-breed analysis, a maximum accuracy of 0.79 was reached with the WssGBLUP<sub>Sum</sub> method and a window size of 40 SNPs and thus, it outperformed WssGBLUP (0.78). For other window sizes (larger or smaller), accuracies with WssGBLUP<sub>Sum</sub> were equal to 0.78. With the WssGBLUP<sub>Max</sub> method, accuracies ranged from 0.77 for a window of two consecutive SNPs to 0.78 for windows of 5, 10, 20, 40, 80, 100, 150, 200 and 250 consecutive SNPs. In comparison, genomic evaluations with

**Table 3 Validation correlations for 351 validation males in the multi-breed population using different WssGBLUP and different window sizes of non-overlapping SNPs**

| Method                  | Size of non-overlapping SNP windows |      |         |                       |
|-------------------------|-------------------------------------|------|---------|-----------------------|
|                         | 1                                   | 2    | 5/10/20 | 40/80/100/150/200/250 |
| WssGBLUP <sup>a</sup>   | 0.79                                |      |         |                       |
| WSSGBLUP <sub>Sum</sub> | 0.79                                | 0.80 | 0.81    |                       |
| WSSGBLUP <sub>Max</sub> | 0.79                                | 0.80 | 0.80    |                       |

<sup>a</sup> Each SNP has its own weight (WssGBLUP standard)

**Table 4 Validation correlations for 205 validation males in the Alpine breed using different WssGBLUP and different window sizes of non-overlapping SNPs**

| Method                  | Size of non-overlapping SNP windows |      |         |      |                    |
|-------------------------|-------------------------------------|------|---------|------|--------------------|
|                         | 1                                   | 2    | 5/10/20 | 40   | 80/100/150/200/250 |
| WssGBLUP <sup>a</sup>   | 0.78                                |      |         |      |                    |
| WSSGBLUP <sub>Sum</sub> | 0.78                                | 0.78 | 0.79    | 0.78 |                    |
| WSSGBLUP <sub>Max</sub> | 0.77                                | 0.78 | 0.78    | 0.78 |                    |

<sup>a</sup> Each SNP has its own weight (WssGBLUP standard)

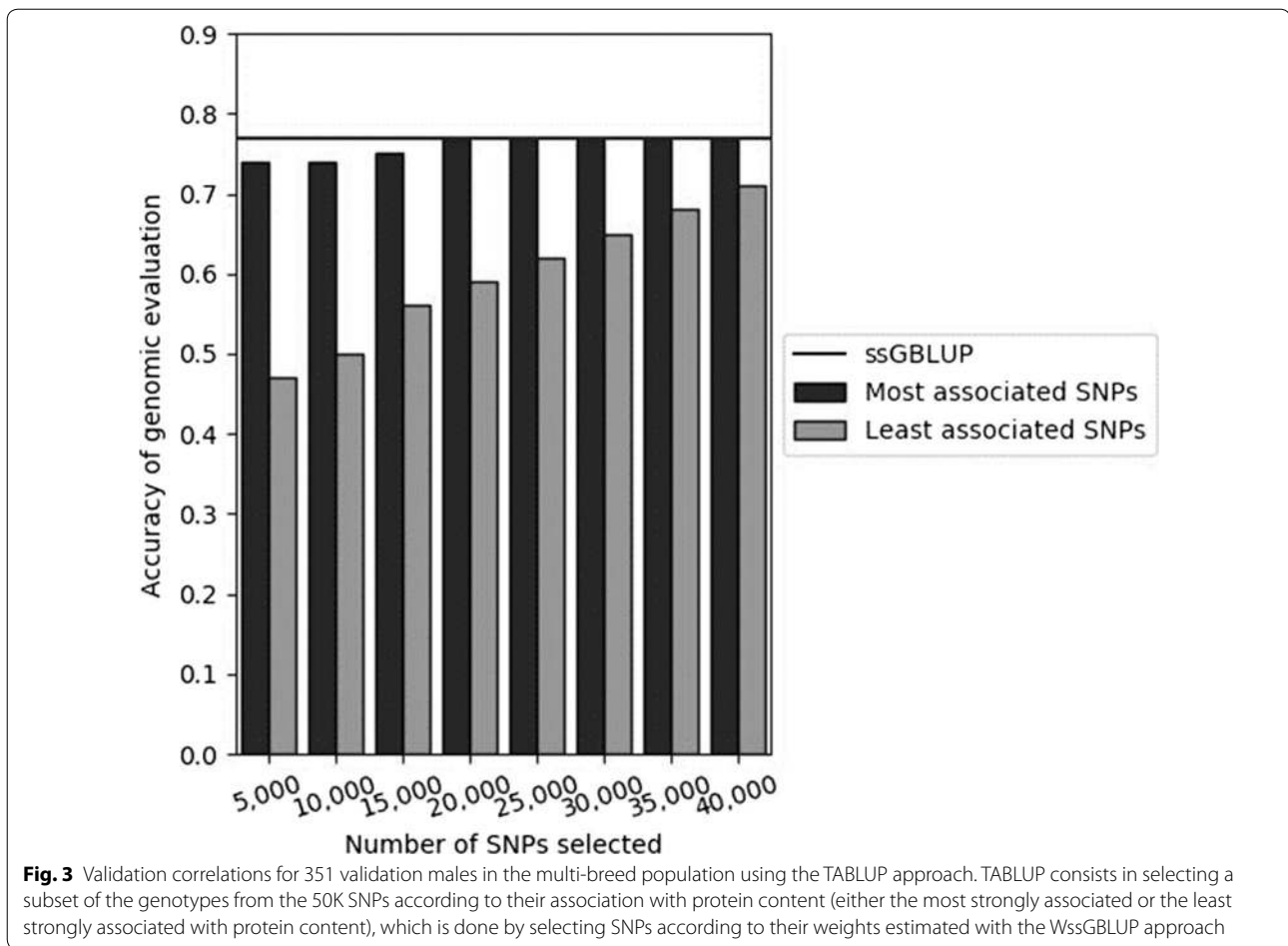
WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> were more accurate than with ssGBLUP (0.76). In the Saanen within-breed analysis, accuracies of 0.78 were reached with WssGBLUP<sub>Sum</sub> for windows of 40, 80, 100, 150, 200 and 250 consecutive SNPs, and with WssGBLUP<sub>Max</sub> for windows of 80 and 100 consecutive SNPs. WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> outperformed WssGBLUP (0.77) or even ssGBLUP (0.73). Accuracies of 0.77 were obtained with WssGBLUP<sub>Sum</sub> for windows of 2, 5, 10 and 20 consecutive SNPs and with WssGBLUP<sub>Max</sub> for windows of 2, 5, 10, 20, 40, 150, 200 and 250 consecutive SNPs.

WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> slightly improved the accuracy of genomic predictions for protein content in French dairy goats compared to WssGBLUP. Similar results were observed by Zhang et al. [24] with WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub> compared to WssGBLUP on simulated data for five QTL. Zhang et al. [27] presented their results for a window size of 20 consecutive SNPs because when they used windows with more than 20 SNPs, accuracies decreased when many QTL affected a trait. This is due to most of the weight being assigned to the windows with large SNP effects and less weight to those with small SNP effects, which may introduce bias in the estimates. For the populations in our study, accuracies varied little with window size. However, 20 consecutive SNPs were not sufficient to reach the highest accuracies and 40 consecutive SNPs were more appropriate. Thus, for a trait that is influenced by few QTL, WssGBLUP<sub>Max</sub> or WssGBLUP<sub>Sum</sub> were more

**Table 5 Validation correlations for 146 validation males in the Saanen breed using different WssGBLUP and different window sizes of non-overlapping SNPs**

| Method                  | Size of non-overlapping SNP window |           |      |        |             |
|-------------------------|------------------------------------|-----------|------|--------|-------------|
|                         | 1                                  | 2/5/10/20 | 40   | 80/100 | 150/200/250 |
| WssGBLUP <sup>a</sup>   | 0.77                               |           |      |        |             |
| WSSGBLUP <sub>Sum</sub> | 0.77                               | 0.78      | 0.78 | 0.78   |             |
| WSSGBLUP <sub>Max</sub> | 0.77                               | 0.77      | 0.78 | 0.77   |             |

<sup>a</sup> Each SNP has its own weight (WssGBLUP standard)



efficient to capture clear signals from QTL compared to WssGBLUP with one weight per SNP.

#### TABLUP method

To validate that ssGBLUP does capture the  $\alpha_{s1}$  *casein* gene information, we used TABLUP that consists in selecting a subset of SNPs for constructing the **G** matrix, i.e. we selected the SNPs that were the most or the least strongly associated with protein content. Figure 3 shows the accuracies obtained with ssGBLUP and TABLUP for the multi-breed population according to the number of SNPs conserved (5000 to 40,000 SNPs) to construct the **G** matrix. Since results for both Alpine and Saanen breeds were similar to those for the multi-breed population, they are not shown.

First, for the SNPs that were the most strongly associated with protein content, TABLUP with only 5000 such SNPs led to a high accuracy of genomic prediction (0.74), which is close to that obtained with ssGBLUP (0.77). TABLUP reached the 0.77 accuracy of ssGBLUP with 20,000 such SNPs, which were distributed across the

whole genome with on average 42% of the SNPs on each chromosome being retained and 54% on chromosome 6. This indicates that SNPs around the  $\alpha_{s1}$  *casein* gene have been more selected than the others. Increasing the number of SNPs from 20,000 to 40,000, did not increase the accuracy furthermore. Conversely, for the SNPs that were the least strongly associated with protein content, TABLUP with 5000 such SNPs led to a very low accuracy (0.47) and increasing their number to 40,000 led to an increase in accuracy of 24 percent points (0.47 with 5000 SNPs and 0.71 with 40,000 SNPs) but accuracy remained significantly lower than that obtained by using the whole 50K SNP BeadChip (0.71 against 0.77).

Using different subsets of SNPs and the BayesA model, VanRaden et al. [44] compared accuracies of genomic predictions in Holstein breed cattle for 33 traits. They used 60K and high-density (HD) SNP panels, and added specific SNPs selected from whole-genome sequence data, which were SNPs based on their annotation (located on exons, splicing sites, indels, 2 kb upstream, 1 kb downstream, untranslated regions, SNPs with large effects). They showed that the highest accuracies were

obtained with the scenario that used 60K SNPs plus the top 1000 SNPs for all 33 traits. Increasing the number of SNPs (using the HD SNP panel for example) did not increase the accuracy of genomic predictions. However, adding selected SNPs from whole-genome sequence to a medium-density SNP BeadChip improved GEBV accuracies. These results agree with those that we obtained with the TABLUP method. In the near future, when whole-genome caprine sequence data become available, it will be possible to select sequence-based variants and add them to the 50K SNP data in the genomic evaluation model, which will improve the accuracy of genomic predictions in these species.

We undertook additional analyses (results not shown) in which SNPs were removed chromosome-wise with the ssGBLUP, WssGBLUP and gene content methods. The same accuracies were observed, regardless of the chromosome from which the SNPs were removed, except for chromosome 6 for ssGBLUP (0.77), WssGBLUP (0.79) and gene content (0.76). When SNPs from chromosome 6 were removed, accuracies dropped to 0.70 for ssGBLUP, 0.66 for WssGBLUP and 0.74 for gene content. However, the loss in accuracy with gene content was smaller than with ssGBLUP and WssGBLUP, i.e. using genotypes for the  $\alpha_{s1}$  *casein* gene and SNPs from 28 chromosomes (except chromosome 6) is quite similar to using the 50K SNP chip. The missing genotypes from the 50K SNP chip (i.e. the SNPs on chromosome 6) did not add much information compared to the information contained by the genotypes for the  $\alpha_{s1}$  *casein*. Results of TABLUP and chromosome-wise removal of SNPs showed that a part of the effect of the  $\alpha_{s1}$  *casein* gene was retained by the ssGBLUP method, which basically does not include information on causal mutations. These results can be explained by the high coverage of SNPs on chromosome 6 around the  $\alpha_{s1}$  *casein* gene.

## Conclusions

Our aim was to investigate different genomic evaluation methods (using  $\alpha_{s1}$  *casein* genotypes and/or 50K SNP information) to integrate information on the  $\alpha_{s1}$  *casein* gene in genomic evaluations of dairy goats. Using the trait-specific marker-derived relationship matrix did not improve accuracy of genomic evaluation, which was the same as that obtained by ssGBLUP with a selection of the 20,000 most important SNPs for protein content. With the gene content method, accuracies of genomic evaluation were not improved compared to ssGBLUP, which is probably due to the  $\alpha_{s1}$  *casein* gene having many alleles and to the small number of genotyped animals. Putting more weight on SNPs with larger effects improved

accuracies of genomic evaluation using WssGBLUP, WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub>. For WssGBLUP<sub>Max</sub> and WssGBLUP<sub>Sum</sub>, accuracies were highest when a common weight was applied to non-overlapping windows of 40 SNPs. Gains in accuracies reached +12 percent points for the Saanen, +9 percent points for the multi-breed and +8 percent points for the Alpine populations compared to a pedigree-based BLUP evaluation. WssGBLUP using common weights for SNPs within non-overlapping windows is efficient if the trait is influenced by few QTL and the true number of QTL is not known. WssGBLUP also combines fast computing and simplicity, and requires ssGBLUP to be run only twice.

## Authors' contributions

MT performed the analysis and wrote the paper. MT, CRG and HL interpreted the results. CRG and HL revised and improved the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This study would not have been possible without the goat SNP50 BeadChip developed by the International Goat Genome Consortium (IGGC): [www.goatgenome.org](http://www.goatgenome.org). The authors thank Ignacy Misztal (University of Georgia, USA) for the blup90iod2 program. We sincerely thank the two anonymous reviewers and Julius Van der Werf (editors for Genetics Selection Evolution) whose comments/suggestions helped improve and clarify this manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because they were partially produced by private professional partnerships.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

All SNP genotyping was done according to the French National Guidelines for the care and use of animals for research.

## Funding

The authors thank the French Genovicap and Phenofinlait programs (ANR, Apis-Gène, CASDAR, FranceAgriMer, France Génétique Elevage, French Ministry of Agriculture Agrifood, and Forestry) and the European 3SR project, which funded part of this work. The first author also received financial support from the Midi-Pyrénées region and the French National Institute for Agricultural Research (INRA) SELGEN program (INCoMINGS).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 8 September 2017 Accepted: 30 May 2018

Published online: 15 June 2018

## References

- Boichard D, Guillaume F, Baur A, Croiseau P, Rossignol MN, Boscher MY, et al. Genomic selection in French dairy cattle. *Anim Prod Sci*. 2012;52:115–20.

2. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92:433–43.
3. Baloché G, Legarra A, Sallé G, Larroque H, Astruc J-M, Robert-Granié C, et al. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J Dairy Sci*. 2014;97:1107–16.
4. Duchemin SI, Colombani C, Legarra A, Baloché G, Larroque H, Astruc J-M, et al. Genomic selection in the French Lacaune dairy sheep breed. *J Dairy Sci*. 2012;95:2723–33.
5. Brito LF, Clarke SM, McEwan JC, Miller SP, Pickering NK, Bain WE, et al. 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*. 2017;18:7.
6. Auvray B, McEwan JC, Newman S. a. N, Lee M, Dodds KG. Genomic prediction of breeding values in the New Zealand sheep industry using a 50 K SNP chip. *J Anim Sci*. 2014;92:4375–89.
7. Mucha S, Mrode R, MacLaren-Lee I, Coffey M, Conington J. Estimation of genomic breeding values for milk yield in UK dairy goats. *J Dairy Sci*. 2015;98:8201–8.
8. Carillier C, Larroque H, Robert-Granié C. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. *Genet Sel Evol*. 2014;46:67.
9. Carillier C, Larroque H, Palhière I, Clément V, Rupp R, Robert-Granié C. A first step toward genomic selection in the multi-breed French dairy goat population. *J Dairy Sci*. 2013;96:7294–305.
10. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
11. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*. 2010;42:2.
12. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 2009;92:4656–63.
13. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score 1. *J Dairy Sci*. 2010;93:743–52.
14. Goddard ME, Hayes BJ, Meuwissen THE. Using the genomic relationship matrix to predict the accuracy of genomic selection: predict the accuracy of genomic selection. *J Anim Breed Genet*. 2011;128:409–21.
15. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
16. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 2008;177:2389–97.
17. Daetwyler HD, Swan AA, van der Werf JH, Hayes BJ. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genet Sel Evol*. 2012;44:33.
18. Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, et al. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet Sel Evol*. 2017;49:17.
19. Andonov S, Lourenco DAL, Fragomeni BO, Masuda Y, Pocrnic I, Tsuruta S, et al. Accuracy of breeding values in small genotyped populations using different sources of external information—a simulation study. *J Dairy Sci*. 2017;100:395–401.
20. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3:e3395.
21. Viana JMS, Piepho H-P, Silva FF. Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations. *Sci Agric*. 2017;74:41–50.
22. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136:245–57.
23. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)*. 2012;94:73–83.
24. Zhang X, Lourenco D, Aguilar I, Legarra A, Misztal I. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. *Front Genet*. 2016;7:151.
25. Strandén I, Garrick DJ. Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci*. 2009;92:2971–5.
26. Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*. 2010;88:544–51.
27. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform*. 2011;12:186.
28. Gianola D. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*. 2013;194:573–96.
29. Vitezica ZG, Aguilar I, Misztal I, Legarra A. Bias in genomic predictions for populations under selection. *Genet Res (Camb)*. 2011;93:357–66.
30. Zhang Z, Ding X, Liu J, de Koning D-J, Zhang Q. Genomic selection for QTL-MAS data using a trait-specific relationship matrix. *BMC Proc*. 2011;5:S15.
31. Gengler N, Mayeres P, Szydlowski M. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal*. 2007;1:21–8.
32. Legarra A, Vitezica ZG. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. *Genet Sel Evol*. 2015;47:89.
33. Carillier-Jacquin C, Larroque H, Robert-Granié C. Including as1 casein gene information in genomic evaluations of French dairy goats. *Genet Sel Evol*. 2016;48:54.
34. Martin P, Palhière I, Maroteau C, Bardou P, Canale-Tabet K, Sarry J, et al. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Sci Rep*. 2017;7:1872.
35. Grosclaude F, Mahé M-F, Brignon G, Di Stasio L, Jeunet R. A Mendelian polymorphism underlying quantitative variations of goat as1-casein. *Genet Sel Evol*. 1987;19:399–412.
36. Larroque H, Astruc JM, Barbat A, Barillet F, Boichard D, Bonaiti B, et al. National genetic evaluations in dairy sheep and goats in France. In: Proceedings of the 62nd annual meeting of the European Federation of Animal Science: 29 August–2 September 2011; Stavanger; 2011.
37. Tosser-Klopp G, Bardou P, Cabau C, Eggen A, Faraut T, Heuven H, et al. Goat genome assembly, availability of an international 50 K SNP chip and RH panel: an update of the International Goat Genome Consortium projects. In: Proceedings of the International Plant and Animal Genome Conference XX: 14–18 January 2012; San Diego; 2012.
38. Misztal I, Tsuruta S, Strabel T, Auvrey B, Druet T, Lee DH. BLUPF90 and related programs. In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production: 19–23 August 20; Montpellier; 2002.
39. VanRaden PM, Wiggans GR. Derivation, calculation, and use of national animal model information. *J Dairy Sci*. 1991;74:2737–46.
40. Williams EJ. The comparison of regression variables. *J R Stat Soc Ser B Methodol*. 1959;21:396–9.
41. Bélichon S, Manfredi E, Piacère A. Genetic parameters of dairy traits in the Alpine and Saanen goat breeds. *Genet Sel Evol*. 1999;31:529–34.
42. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited review: reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*. 2009;92:16–24.
43. Su G, Christensen OF, Janss L, Lund MS. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *J Dairy Sci*. 2014;97:6547–59.
44. VanRaden PM, Tooker ME, O'Connell JR, Cole JB, Bickhart DM. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet Sel Evol*. 2017;49:32.