Using sequence variants of a QTL region improves the accuracy of genomic evaluation in French Saanen goats

Estelle Talouarn, Marc Teissier, Philippe Bardou, Hélène Larroque, Virginie Clément, Isabelle Palhière, Gwenola Tosser-Klopp, Rachel Rupp, Christèle Robert-Granié

J. Dairy Sci. 104:588–601
https://doi.org/10.3168/jds.2020-18837

Using sequence variants of a QTL region improves the accuracy of genomic evaluation in French Saanen goats

Estelle Talouarn,[1]* Marc Teissier,[1] Philippe Bardou,[2] Hélène Larroque,[1] Virginie Clément,[3] Isabelle Palhière,[1] Gwenola Tosser-Klopp,[1] Rachel Rupp,[1] and Christèle Robert-Granié[1]
[1]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326 Castanet-Tolosan, France
[2]Sigenae, INRAE, F-31326 Castanet-Tolosan, France
[3]Institut de l'Elevage, F-31326 Castanet-Tolosan, France

ABSTRACT

The enhanced availability of sequence data in livestock provides an opportunity for more accurate predictions in routine genomic evaluations. Such evaluations would therefore no longer rely only on the linkage disequilibrium between a chip marker and the causal mutation. The objective of this study was to assess the usefulness of sequence data in Saanen goats (n = 33) to better capture a quantitative trait locus (QTL) on chromosome 19 (CHI19) and improve the accuracy of predictions for 3 milk production traits, 5 type traits, and somatic cell scores. All 1,207 50K genotypes were imputed to the sequence level. Four scenarios, each using a subset of CHI19 imputed variants, were then tested. Sequence-derived information included all CHI19 variants (529,576), all variants in the QTL region (22,269), 178 variants selected in the QTL region and added to an updated chip, or 178 randomly selected variants on CHI19. Two genomic evaluation models were applied: single-step genomic BLUP and weighted single-step genomic BLUP. All scenarios were compared with single-step genomic BLUP using 50K genotypes. Best overall results were obtained using single-step genomic BLUP on 50K genotypes completed with all variants in the QTL region of chromosome 19 (6.2% average increase in accuracy for 9 traits) with the highest accuracy gain for fat yield (17.9%), significant increases for milk (13.7%) and protein yields (12.5%), and type traits associated with CHI19. Despite its association with the QTL region of chromosome 19, the somatic cell score showed decreased accuracy in every alternative scenario. Using all CHI19 variants led to an overall decrease of 4.8% in prediction accuracy. The updated chip was efficient and improved genomic evaluations by 3.1 to 6.4% on average, depending on the scenario. Indeed, information from only a few carefully selected variants increased accuracies for traits of interest when used in a single-step genomic BLUP model. In conclusion, using QTL region variants imputed from sequence data in single-step genomic evaluations represents a promising perspective for such evaluations in dairy goats. Furthermore, using only a limited number of selected variants in QTL regions, as available on SNP chip updates, significantly increases the accuracy for QTL-associated traits without deteriorating the evaluation accuracy for other traits. The latter approach is interesting, as it avoids time-consuming imputation and data formatting processes and provides reliable genotypes.
Key words: genomic evaluations, sequence, Saanen, dairy goats

INTRODUCTION

The recent decrease in sequencing costs has made it possible to sequence large numbers of individuals in livestock species. Including sequence data in genomic evaluations is interesting because it might improve the persistency of genomic prediction accuracy (Hayes et al., 2014). Indeed, such evaluations would no longer rely simply on the linkage disequilibrium (**LD**) between a chip marker and the causal mutation of a trait. In sheep (Moghaddar et al., 2018) and dairy cattle (Hayes et al., 2014), the gain in accuracy when sequence data were included in the kinship matrix calculation ranged from 1.4 to 2.6% compared with genomic evaluations performed on 50K genotypes and reached up to 2% when compared with high-density (**HD**) genotypes in dairy cows.

The VarGoats program made available over 1,000 sequences of the *Capra* genus from 125 breeds (http://www.goatgenome.org/vargoats.html). Talouarn et al. (2020) investigated the feasibility of imputation from 50K-chip information using the sequence data available for 33 sequenced French Saanen individuals. Acceptable imputation accuracy was achieved within-breed with mean allele and genotype concordance rates of 0.86 and 0.74 respectively in the Saanen breed. The imputation

Received May 4, 2020.
Accepted August 11, 2020.
*Corresponding author: estelle.talouarn@inrae.fr

to sequence data also fine mapped a previously identified QTL in Saanen goats and led to the identification of new signals in both Alpine and Saanen breeds (Talouarn et al., 2020). The QTL region of chromosome 19 (CHI19) explains between 5 and 10% of the total additive genetic variance of SCS and type traits (Martin et al., 2018). The refined information for CHI19 provided by sequence data might be of interest when performing genomic evaluations in Saanen goats because 6 of the 11 traits included in the indices are associated with the QTL (Martin et al., 2017, 2018; Talouarn et al., 2020).

France is a leading European country in goat milk production. Selection strategies aim to improve cheese making. A synthetic production index has been established using production traits of milk, protein, and fat yields and protein and fat contents. Type traits, such as fore udder, teat orientation, udder floor position, udder profile, and rear udder attachment, are now also included in a morphology index. Indices for both production and type traits are combined in a synthetic index which differs between Alpine and Saanen breeds (Clément et al., 2006; Larroque et al., 2011). Since 2013, udder health is considered for AI buck selection using SCC (Virginie Clément, Institut de l'Elevage, Castanet-Tolosan, France, personal communications). Following the introduction in 2011 of the 50K genotyping chip (Illumina GoatSNP50 BeadChip, Illumina Inc., San Diego, CA; Tosser-Klopp et al., 2014), the feasibility of genomic evaluations in French dairy goats has been studied (Carillier et al., 2013). Such evaluations were officially implemented in French Alpine and Saanen in 2018 using a single-step genomic BLUP model (**ssGBLUP**; Legarra et al., 2009). Current perspectives for improving genomic evaluations rely on finding better ways of integrating genotype information. Studies were therefore performed to include major gene information in the evaluations using weighted ssGBLUP (**WssGBLUP**; Teissier et al., 2018, 2019). The SNP close to the casein α-S1 gene had higher weights and led to substantial gains in accuracy for the genomic estimated breeding values (**GEBV**) of protein content in Alpine and Saanen goats (Teissier et al., 2018). As mentioned, other studies identified a large QTL region on CHI19 for udder type, udder health, and milk production traits (Martin et al., 2017, 2018; Mucha et al., 2018) in Saanen and mixed-breed goats. The WssGBLUP led to gains in accuracy ranging from 2 to 14% in Saanen goats for traits associated with the QTL region of CHI19 (Teissier et al., 2019).

Here we describe the first exploratory study of genomic evaluations using information extracted from sequence data in French Saanen goats. Our objective was to take advantage of available sequence data (n = 33) to include refined CHI19 information in French Saanen genomic evaluations for 9 traits: milk yield, fat yield, protein yield, fore udder, teat orientation, udder floor position, udder profile, rear udder attachment, and SCS. We investigated the relevance of including sequence variants in routine genomic evaluations for these 9 traits. We focused on identifying the method that would maximize both the accuracy of prediction and the computation efficiency.

MATERIALS AND METHODS

This study did not require ethical approval because no experiments on animals were necessary (samples originated from other studies).

Animals, Phenotypes, and 50K Genotypes

Details on phenotypes and 50K-genotype quality checks are described in Teissier et al. (2018). The milk performance data set was provided by the French national milk records system and the udder traits records by the breeding company Capgenes (Mignaloux-Beauvoir, France). Phenotypes for milk production were considered over the whole lactation period: 250-d milk yield (**MY**, in kg), 250-d protein and fat yields (**PY** and **FY**, respectively, both in kg), and 250-d somatic cell score (**LSCS**). The type traits were fore udder (**FU**), teat orientation (**TO**), udder floor position (**UFP**), udder profile (**UP**), and rear udder attachment (**RUA**). Type traits were only measured once per animal, mainly during their first lactation and sometimes in their second. Phenotypes, pedigree data, and genotypes were obtained from the official genetic evaluation of January 2016 (Larroque et al., 2011). Phenotype data were retained only for French Saanen goats born between 1980 and 2017. The final data set is described for each trait in Table 1.

The pedigree consisted of 2,177,617 individuals and was complemented by defining unknown parent groups. Sixteen groups were defined according to the year of birth of the descendants: before 1975, between 1975 and 1980, between 1980 and 1983, and then every 2 years. Males and females were pooled together in unknown parent groups because there were few animals with unknown dams.

Genotypes were acquired with the Illumina GoatSNP50 BeadChip. A total of 1,207 genotypes (394 males, 813 females) were retained for French Saanen goats after the quality check step. The quality check step was previously described in Talouarn et al. (2020). It implies removing all individuals with a call rate below 95% or showing pedigree inconsistency. The SNP quality control was based on the following inclusion criteria: call rate above 99%, minor allele frequency above 1%,

and Hardy-Weinberg P-value above 10^{-6}. After quality control, 47,147 SNP were retained, including 1,143 SNP on chromosome 19.

Sequence-Derived Information

Quality Check of Sequence Data and Imputation. Sequence data are described in Talouarn et al. (2020). The sequence data were retrieved from the VarGoats project (http://www.goatgenome.org/vargoats.html). The 37 French Saanen individuals came from VarGoats child projects PRJEB37276, PRJEB37276, and NextGen PRJEB5900. The final data set comprised 33 French Saanen goats (31 males and 2 females); 4 individuals were removed as their mean coverage was below 5. All of these were also genotyped with the Illumina GoatSNP50 BeadChip.

A wide QTL region was previously identified in Saanen goats on CHI19 between 24.72 and 28.38 Mb (Martin et al., 2018; Mucha et al., 2018; Talouarn et al., 2020). This region is associated with production traits (milk, fat, and protein yields), stature, udder type and health, as well as semen volume. Sequence quality check and soft filtering processes were described in detail by Talouarn et al. (2020) and resulted in keeping 23,337,436 variants, including 539,476 variants on chromosome 19 and 22,269 variants between 24.72 and 28.38 Mb. Before imputing available 50K genotypes, imputation was necessary to fill in the gaps of the sequenced panel. A combination of AlphaImpute (v.1.9; Hickey et al., 2012) and FImpute (v.3.0; Sargolzaei et al., 2014) was used because it gave higher concordance rates between 50K genotypes and sequence than using only one software and minimized computation time. The mean concordance rate between 50K-genotypes and sequence data, calculated on the 33 sequenced Saanen goats, was 98.43% (±1.35) after filtering and imputation. No missing genotypes remained for sequenced Saanen after these steps.

Finally, imputation of the 1,207 50K-genotypes was performed using pedigree information on chromosome 19. Mean genotype and allele concordance rates were estimated at 71.8 and 84.3%, respectively, in the Saanen breed.

Illumina GoatSNP50 BeadChip Update. The Illumina GoatSNP50 BeadChip is currently being updated with a set of about 6,832 probes (including duplicates) to be validated, resulting in the Goat IGGC 65k v2 chip. The region between 23 and 30 Mb on chromosome 19 was densified to better capture the previously identified QTL. The 178 variants in the QTL region added to the chip were chosen based on association analysis using sequence information (E. Talouarn, unpublished results). Variants were selected using the following criteria: (1) P-value in the association analysis, (2) number of traits significantly linked to the variant, (3) minor allele frequency profile in French Saanen goats, (4) spacing within the signal, and (5) distance to SNP of the current version of the chip. The 178 positions were extracted from imputed sequence data to quantify the effect of the increased representation of chromosome 19 (+178 variants) on the accuracy of genomic evaluations. To determine whether gains in precision were only due to the increased SNP density, another 178 variants were also randomly selected on chromosome 19 using PLINK software (Purcell et al., 2007). Among them, only 11 were located in the QTL region of chromosome 19.

Evaluation Methods

Single-Step GBLUP. For ssGBLUP, the model consisted of the following:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}, \quad [1a]$$

where $\mathbf{y}$ is the vector of performances for the studied trait, $\boldsymbol{\beta}$ is the vector of fixed effects defined as in the

Table 1. Summary statistics for the traits used in the genetic evaluations of French Saanen goats (born between 1980 and 2017)[1]

Trait[2]	Number of lactations	Number of females with phenotypes	Minimum	Mean	SD	Maximum
MY (kg)	3,470,255	1,242,020	34.51	837.05	266.03	2,615.04
PY (kg)	3,470,255	1,274,581	1.27	25.05	8.08	84.64
FY (kg)	3,470,255	1,271,383	1.09	28.06	10.01	111.92
LSCS	1,449,698	705,753	−0.58	8.82	1.34	13.57
FU	—	160,086	1	3.29	1.16	9
TO	—	160,086	1	4.03	0.86	9
UFP	—	160,086	1	6.23	1.14	9
UP	—	160,086	1	6.28	1.34	9
RUA	—	160,086	1	4.83	1.67	9

[1]Measurements for type traits were performed once in each animal's lifetime.

[2]MY = milk yield; PY = protein yield; FY = fat yield; LSCS = 250-d SCS; FU = fore udder; TO = teat orientation; UFP = udder floor position; UP = udder profile; RUA = rear udder attachment.

official genomic evaluations. For production traits and LSCS, fixed effects were defined for each year and parity and were: herd, age at kidding × 4 geographic regions, month of kidding × 4 regions, length of dry period × 4 regions; **u** is a vector of random additive genetic effects assumed to be normally distributed $N\left(\mathbf{0},\mathbf{H}\sigma_u^2\right)$, where σ_u^2 is the variance of u; **p** is the vector of random permanent environmental effects assumed to be normally distributed $N\left(\mathbf{0},\mathbf{I}\sigma_p^2\right)$, where **I** is the identity matrix; **e** is a vector of random residuals also normally distributed $N\left(\mathbf{0},\mathbf{I}\sigma_e^2\right)$; **X** is the incidence matrix relating phenotypes and the fixed effects; **Z** and **W** are the design matrices linking phenotypes to genetic and permanent environmental effects, respectively. The **H** matrix is the genetic relationship matrix that integrates both genotype and pedigree information implemented as in Legarra et al. (2009):

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

where **A** is a pedigree-based relationship matrix with indices 1 for ungenotyped animals and 2 for genotyped animals, and **G** is the genomic relationship matrix derived as in Christensen and Lund (2010):

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

where m is the number of variants, p_i is the estimated allele frequency at the locus i, and **M** is a centered matrix of variant genotypes.

For udder traits, which were measured only once in the life of goats, the genomic evaluation model did not include a permanent environmental effect, and was as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \qquad [1b]$$

where **y**, **u**, and **e** are the same as previously stated; **β** includes 3 combined fixed effects: herd × parity × year, age at scoring × year, and days in milk at scoring × year. Variance components were estimated by using the restricted maximum likelihood method in the remlf90 software (Misztal et al., 2002) and ssGBLUP analyses were performed with the blup90iod2 software (Misztal et al., 2002).

Weighted Single-Step GBLUP. The same model was used to perform WssGBLUP. The WssGBLUP was applied to each trait studied using blupf90 family software (Misztal et al., 2002). The SNP effects and SNP weights were estimated using postGSf90 software (Misztal et al., 2002). Our objective was to take advantage of the sequence information for chromosome 19 to increase the accuracy of genomic evaluations for the traits related to the QTL region. The WssGBLUP is based on iterative ssGBLUP in which weights for SNP variances are used to form the genomic relationship matrix **G**. This method is used to give more weight to causal mutations, variants that are in high LD with a causal mutation or variants within a QTL region with a relatively large effect. The genomic relationship matrix **G** was built differently. The solutions for genomic breeding values from ssGBLUP can be decomposed into SNP effects as follows (Wang et al., 2014):

$$\hat{\mathbf{a}} = \mathbf{DM}'(\mathbf{MDM}')^{-1}\hat{\mathbf{u}}_{\mathbf{g}},$$

where $\hat{\mathbf{a}}$ is a vector of variant effects, **D** is a diagonal matrix of weights (set to 1 in ssGBLUP), **M** is the centered matrix of variant genotypes, and $\hat{\mathbf{u}}_{\mathbf{g}}$ is the vector of GEBV from genotyped animals only. The additive variances of the effect of variant i were estimated as

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

where p_i is the allele frequency of variant 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 SNP) and used as weights in matrix **D** to construct the weighted matrix **G** ($\mathbf{G}^*$) as described by Wang et al. (2014):

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

The GEBV were estimated again with models [1a] and [1b] by considering weights for each SNP via the $\mathbf{G}^*$ matrix included in the **H** matrix. This process was carried out iteratively with weights estimated at each iteration as described by Wang et al. (2012). The following formulas were used as described by Wang et al. (2012):

1. Iteration 1, initialization, $\mathrm{D}_{(1)} = \mathrm{I}$; $\mathbf{G}^* = 0.95 \times \lambda\mathbf{Z}\mathrm{D}_{(1)}\mathbf{Z}' + 0.05 \times \mathbf{A}_{22}$.
2. Calculation of $\mathbf{G}^*$ following the previous formula to obtain the EBV vector $\hat{\mathbf{u}}_{\mathbf{g}}$.
3. Iteration 2 (*it*).

4. Estimation of variant effects $\hat{\mathbf{a}}_{(it)} = \lambda D_{(it-1)}\mathbf{Z}'\mathbf{G}^{*-1}_{(it-1)}\hat{\mathbf{u}}_{g(it-1)}$.
5. Conversion of effects into weights following the formula: $d_i^* = \hat{\mathbf{a}}_i^2 \times 2p_i\left(1-p_i\right)$. Weights are integrated into matrix $\mathbf{D}^*_{(\mathbf{it})}$.
6. Weights normalization $\mathbf{D}_{(\mathbf{it})} = \dfrac{tr\left(D_{(1)}\right)}{tr\left(D^*_{(it)}\right)} \times D^*_{(it)}$; tr s the trace of the $\mathbf{D}$ matrix.
7. Building $\mathbf{G}^*_{(\mathbf{it})}$. $\mathbf{G}^*_{(\mathbf{it})} = 0.95 \times \lambda \mathbf{Z} D_{(it)}\mathbf{Z}' + 0.05 \times \mathbf{A}_{22}$.
8. Launch of a WssGBLUP using the newly calculated $\mathbf{G}^*_{(\mathbf{it})}$ matrix to obtain new EBV.
9. Exit.

In French dairy goats, Teissier et al. (2018) showed that 2 iterations were sufficient to maximize the accuracy. We therefore stopped after 2 iterations in our study. This model will be referred to as standard WssGBLUP.

Weighted Single-Step GBLUP Using Windows

As proposed by Zhang et al. (2016), several other methods can be considered to calculate the weight for variants in the $\mathbf{D}$ matrix. These methods assign the same weight to several consecutive SNP within a chromosomal region. These methods have already been explored by Teissier et al. (2018) in French dairy goats. The best results were obtained when using nonoverlapping windows of 40 consecutive SNP and taking the maximum weight in the window. These weights were calculated based on the weights estimated with WssGBLUP. The highest weight observed in a window was assigned to all SNP within the same window. When adding sequence information for the QTL region on chromosome 19 (22,269 sequence variants between 23 and 30 Mb), the average spacing between variants was much lower than on other chromosomes (2,670 bp ± 13,290 vs. 60,000 bp for the rest of the genome). We decided to take this information into account when building the windows. We tested 2 options: (1) using windows of 40 consecutive variants, or (2) using windows of 2.4 Mb (average spacing of 40 consecutive SNP on the chip).

Tested Scenarios

Several scenarios were tested, based on either ssGBLUP or WssGBLUP methods, and using 4 different sets of new markers in addition to the Illumina GoatSNP50 BeadChip (Table 2). We wanted to see how we could expect the next version of the Illumina GoatSNP50 BeadChip to perform if used for the genomic evaluations (geno_50kv2QTL). As mentioned previously, the 178 additional variants selected in the QTL region of chromosome 19 and imputed in our data set were extracted for that purpose. To assess the relevance of the results, another scenario consisted of including 178 randomly selected variants located over the whole length of chromosome 19 (geno_50kv2_random). In scenarios 3 and 4, all imputed variants from the sequence on chromosome 19 (n = 539,476) (geno_50kseqCHI19) or only those in the QTL region (22,269; geno_50kseqQTL) were chosen, respectively. As our goal was to find the best candidate scenario to improve genomic predictions, all scenarios were systematically compared with single-step genomic evaluation using 50K genotypes (geno_50k).

Single-Step GBLUP

The different scenarios and information used are summarized in Table 2.

Weighted Single-Step GBLUP

Our objective using WssGBLUP was to compare the integration of sequence data for the QTL region of chromosome 19, i.e., geno_50kseqQTL (22,269 sequence variants in the QTL region) or geno_50kv2QTL (178 SNP in the QTL region) with a simple evaluation using 50K genotypes (geno_50k). This was performed using a standard WssGBLUP model.

We also attempted to use windows of consecutive markers ($\mathbf{WssGBLUP}_{\mathbf{windows}}$). The highest weight in the window was assigned to all the SNP in the same window. Because the average spacing in the QTL region was smaller than when solely using 50K genotypes for all 4 approaches, we implemented windows of 2.4 Mb which is the average distance covering 40 SNP with the Illumina GoatSNP50 BeadChip. This window size has proven to be efficient in previous studies using 50K genotypes (Teissier et al., 2018). We retained the alternative using windows of 40 consecutive SNP for comparison purposes. We implemented 4 approaches: (1) WssGBLUP on 50K genotypes building windows of 2.4 Mb over the whole genome, (2) WssGBLUP on 50K genotypes and sequence information for the QTL region (either 178 selected variants or 22,269 sequence variants) building 2.4-Mb windows only on chromosome 19, (3) WssGBLUP on 50K genotypes and sequence information for the QTL region (either 178 selected variants or 22,269 sequence variants) building windows of 40 consecutive variants over the whole genome, and

Table 2. Summary of the scenarios tested using a single-step genomic BLUP model

Name of the scenario	Variants included in the scenario					Total number of variants
	50K markers	178 SNP in chromosome 19 QTL region on the chip update	178 randomly selected SNP on chromosome 19	539,476 sequence variants over the whole chromosome 19[1]	22,269 sequence variants in the QTL region of chromosome 19[1]	
Ref geno_50k	x					47,147
1 geno_50kv2QTL	x	x				47,325
2 geno_50kv2_random	x		x			47,325
3 geno_50kseqCHI19	x			x		586,623
4 geno_50kseqQTL	x				x	69,416

[1]Variants obtained on imputed sequence of 1,207 Saanen individuals.

(4) the same scenario as (3) but building windows of 2.4 Mb.

Accuracy of Genomic Predictions

The reference population used to assess the accuracy of genomic evaluation comprised only genotyped males, even if the genotypes of females were also used in ssGBLUP and WssGBLUP evaluations. This reference population was split into 2 subsets: a training set and a validation set. The training population consisted of 248 sires born between 1998 and 2007 and genotyped with the Illumina GoatSNP50 BeadChip. All the information for these animals (genotype, pedigree with their ancestry and progeny, and phenotypes of their progeny) was retained in the data sets to estimate the GEBV. The validation population consisted of 146 bucks born between 2008 and 2012. All of their progeny with phenotypes were removed from the data set. The GEBV estimated in these conditions and the daughter yield deviation (**DYD**) computed from the official genetic evaluation of January 2016, were compared for the 146 animals in the validation set. The DYD were the average performance values for the daughters corrected for fixed and random environmental effects and half of the merit of their dams. The DYD were weighted by effective daughter contributions as described by VanRaden and Wiggans (1991). The accuracy of genomic predictions was assessed by calculating the Pearson correlation between the GEBV estimated with each model and DYD. To compare the Pearson correlations obtained with the different scenarios and methods, we used the Hotelling-Williams test as implemented in the multilevel R package (Williams, 1959).

RESULTS

All tested scenarios were systematically compared with ssGBLUP using 50K genotypes (geno_50k).

Single-Step GBLUP

Figure 1 compares the accuracy of evaluations with the current version of the chip (geno_50k), the updated version of the chip (178 additional variants, geno_50v2QTL), the 178 randomly selected variants on chromosome 19 (geno_50kv2QTL_random) and evaluations using either sequence data of the QTL region (geno_50kseqQTL) or the imputed variants of the whole chromosome 19. The percentage of variance explained for each trait in the geno_50k, geno_50kv2QTL and geno_50kseqQTL scenarios were calculated per variant in a ssGBLUP model using the option *windows_variance* parameter of BLUPF90.Variance explained by

the QTL region was obtained by summing the variance explained by each SNP and are shown in Figure 2. In the geno_50kseqQTL scenario, adding sequence information with a known chromosomal location led to significant gains ($P < 0.05$) in accuracy (up 6.2% on average compared with geno_50k). The highest gain in accuracy was observed for FY (17.9%), and significant ($P < 0.05$) increases were obtained for MY (13.7%), PY (12.5%), FU (4.5%), UFP (9.0%), and RUA (4.9%). For LSCS, the accuracy decreased by 7.1% (Figure 1). However, when information was added over the whole chromosome (539,476 variants, geno_50kseqCHI19 scenario), the accuracy of the evaluations decreased by 4.8% on average compared with evaluations performed solely on 50K genotypes (geno_50k scenario). Two significant ($P < 0.05$) increases in accuracy were observed in the geno_50kseqCHI19 scenario for MY (7.9%) and RUA (3.6%). The highest decrease was observed for TO with a loss of 14.7% of accuracy, although not significant.

The additional selected variants on the updated version of the chip (geno_50kv2QTL) significantly ($P < 0.05$) increased the accuracy of genomic predictions for all the traits associated with the QTL region of chromosome 19 (FU, UFP, RUA, LSCS, MY, FY) except for PY. The mean gain was 3.4% for all traits considered. The highest gain was observed for FY (an increase of 10.0%), whereas the greatest loss was observed for PY (a decrease of 7.4%). On the other hand, selecting randomly 178 variants on chromosome 19 did not have a significant effect on the accuracy of genomic evaluations for any of the traits evaluated except for FU, which was marginally decreased by 0.1%.

In conclusion, geno_50kseqQTL outperformed all the other scenarios for the traits associated with the QTL region without significantly decreasing the accuracy of evaluation for the other traits. The use of variants located in the QTL region systematically increased the accuracy of predictions compared with the same number of variants spread over the whole chromosome.

WssGBLUP

Table 3 shows the results for WssGBLUP (one weight for each SNP) for the different scenarios tested. When performing WssGBLUP on 50K genotypes (geno_50k), the gain in accuracy was 0.3% on average for all traits, ranging from −14.6% (LSCS; not significant) to +11.5% (MY). Including the markers selected for the chip update (geno_50kv2QTL) led to an average gain in accuracy of 2.9% with a significant ($P < 0.05$) decrease for LSCS (−14.1%) and a significant increase for production traits, especially FY (+15.5%). The mean gain when using 50K genotypes and sequence variants of the QTL region (geno_50kseqQTL) was 2.1% ranging from −15.8% (LSCS) to +15.9% (FY).

In conclusion, WssGBLUP does not significantly improve accuracies compared with ssGBLUP models. This method is only beneficial for production traits which were also improved when using ssGBLUP.

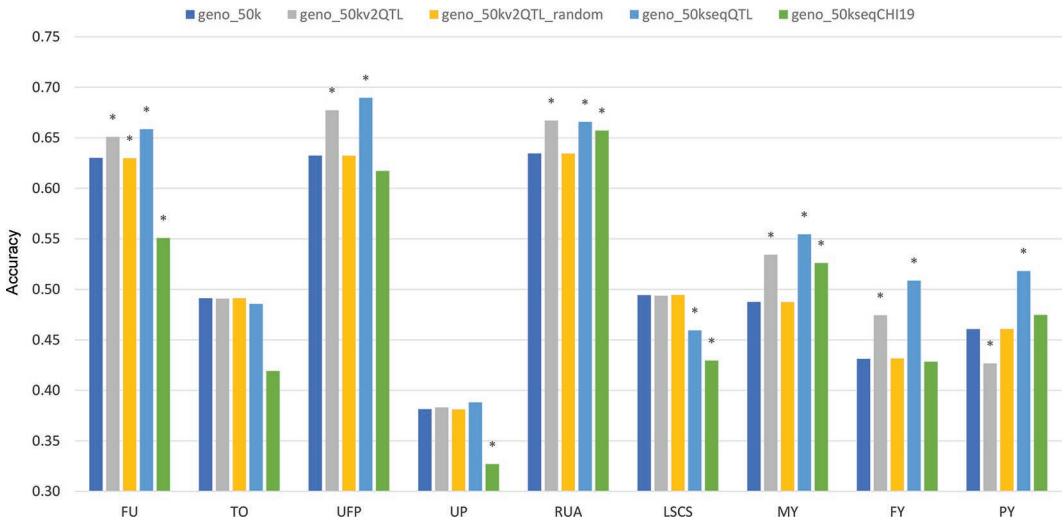

Figure 1. Accuracies of genomic predictions for single-step genomic BLUP (ssGBLUP) model on different scenarios in the 146 validation Saanen individuals. * indicates significant difference from accuracy obtained in a ssGBLUP using 50K genotypes ($P < 0.05$). FU = fore udder; FY = fat yield, LSCS = somatic cell score; MY = milk yield; PY = protein yield; RUA = rear udder attachment; TO = teat orientation; UFP = udder floor position; UP = udder profile.

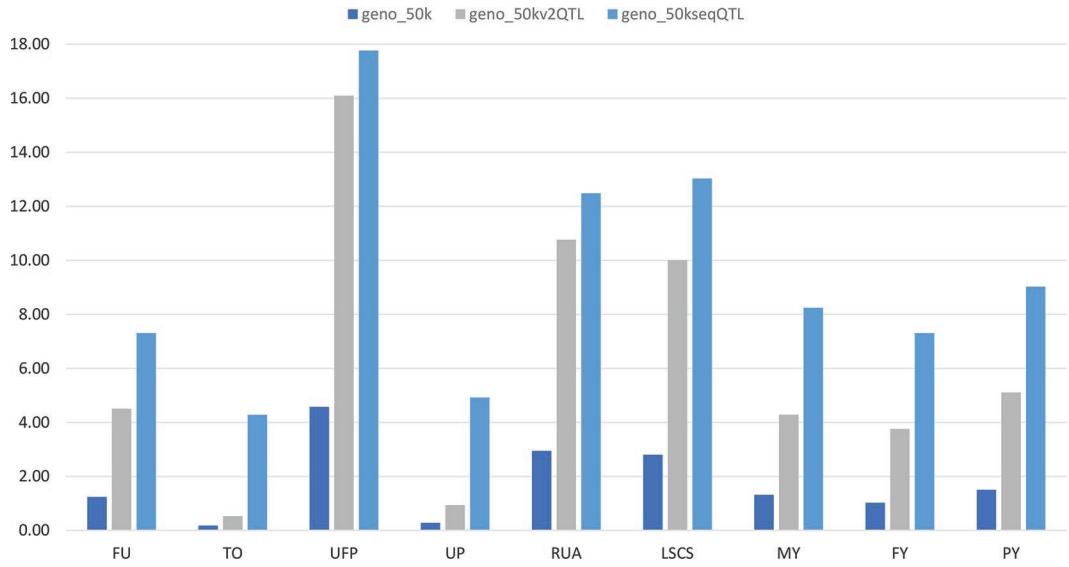

Figure 2. Percentage of the variance explained by the region between 23 and 30Mb of chromosome 19 (CHI19) for each trait in the geno_50k (130 variants), geno_50kv2QTL (308 variants), and geno_50kseqQTL (22,399 variants) scenarios. FU = fore udder; FY = fat yield, LSCS = somatic cell score; MY = milk yield; PY = protein yield; RUA = rear udder attachment; TO = teat orientation; UFP = udder floor position; UP = udder profile.

WssGBLUP$_{windows}$

Every scenario using either 2.4-Mb or 40-SNP windows gave similar results (Table 4). When using the updated version of the chip (geno_50kv2QTL), the mean gain in accuracy was 6.4% and 6.3% for the 2.4-Mb and 40-SNP windows, respectively. In this scenario, the highest gain was observed for MY with 19.0 and 19.2% for the 2.4-Mb and 40-SNP windows, respectively. The greatest losses were observed for LSCS with a significant ($P < 0.05$) decrease of 6.2 and 7.1% for 2.4-Mb and 40-SNP windows, respectively.

Using all variants of the QTL region (geno_50kseqQTL) tended to be slightly less accurate with an average gain of 4.2 and 4.5% for the 2.4-Mb and 40-SNP windows respectively. For this scenario, the highest accuracies were observed for FY in both scenarios with increases comprised between 14.9 and 18.7%. Highest losses were observed for LSCS with a decrease between 9.2 and 10.8%.

When the windows were located on the chromosome 19 only, the accuracy varied marginally with an average decrease of 0.5%. The decrease was particularly significant ($P < 0.05$) for LSCS (12.4%), UP (10.9%), and TO (10.7%). The MY was better predicted with an increase in accuracy of 15.8%.

In conclusion, WssGBLUP$_{windows}$ does not improve genomic evaluations as well as a ssGBLUP model except for production traits for which it results in high gains in accuracy.

Table 3. Accuracy of genomic predictions using standard weighted single-step genomic BLUP (WssGBLUP) in the validation population of 146 Saanen bucks

Trait[1]	geno_50k	geno_50v2QTL	geno_50kseqQTL
FU	0.60*	0.62*	0.60*
TO	0.48*	0.48	0.48
UFP	0.66*	0.66*	0.67*
UP	0.35*	0.35*	0.35*
RUA	0.63	0.64	0.63
LSCS	0.42	0.43*	0.42*
MY	0.54*	0.57*	0.56*
FY	0.47*	0.50*	0.50*
PY	0.50*	0.53*	0.53*

[1]FU = fore udder; FY = fat yield; LSCS = 250-d SCS; MY = milk yield; PY = protein yield; RUA = rear udder attachment; TO = teat orientation; UFP = udder floor position; UP = udder profile.

*Significantly different from the correlation obtained with single-step genomic BLUP using solely 50K genotypes ($P < 0.05$).

DISCUSSION

Variant Selection

As a result of the efforts of the VarGoats Consortium, a large amount of sequence data are now available to the research community. However, sequence data comprise vast numbers of variants (23,337,436 for the whole genome after the filtering process). Careful variant selection is therefore crucial to avoid burdening genomic evaluations with too much information. In this study, we assessed various options for selecting variants. We are aware that the imputation quality is lower than in other species which might lead to a deg-

Table 4. Accuracy of genomic predictions using weighted single-step genomic BLUP (WssGBLUP) with different windows in the validation population of 146 Saanen goats[1]

	geno_50k	geno_50kv2QTL		geno_50kseqQTL		
Item	2.4 Mb (≈40 SNP)	40 SNP	2.4Mb	40 SNP	2.4 Mb	2.4Mb on CHI19 only
FU	0.62*	0.64	0.64	0.63	0.63	0.57*
TO	0.48*	0.48*	0.48*	0.47*	0.47*	0.44*
UFP	0.67*	0.69*	0.68*	0.68*	0.67*	0.66*
UP	0.38	0.38	0.38	0.37	0.37	0.34*
RUA	0.65*	0.67*	0.67*	0.66*	0.65	0.65
LSCS	0.48	0.46*	0.46*	0.45*	0.44*	0.43*
MY	0.53*	0.58*	0.58*	0.56*	0.56*	0.54*
FY	0.44*	0.50*	0.50*	0.50*	0.51*	0.50*
PY	0.49*	0.54*	0.54*	0.52*	0.52*	0.49

[1]FU = fore udder; FY = fat yield; LSCS = 250-d SCS; MY = milk yield; PY = protein yield; RUA = rear udder attachment; TO = teat orientation; UFP = udder floor position; UP = udder profile; CHI19 = chromosome 19.

*Significantly different from the correlation obtained with single-step genomic BLUP using solely 50K genotypes ($P < 0.05$).

radation of predictions as suggested by Perez-Enciso et al. (2015). However, the genomic predictions were improved in our study. Further investigations will be needed once the update of the current chip will be validated to confirm the improvement we observed with more reliable genotypes. Indeed, it has previously been shown that when all variants on chromosome 19 (539,476 variants) are included in ssGBLUP, the genomic evaluations tend to be less accurate. Consistent with this finding, we found that adding sequence data for the whole chromosome 19 introduced noise in the evaluations compared with genomic evaluations using chip data only. Figure 3 shows the variant effects for the FU trait in the different scenarios. Among the traits associated with the QTL region, FU is the most affected in the geno_50kseqCHI19 scenario (12.6% decrease in accuracy). The introduction of sequence data of the whole CHI19 (geno_50kseqCHI19) led to a disruption of estimated effects on the whole genome (Figure 3). The distribution of the variant effects is completely modified including effects of the 50K SNP variants. The latter are greater than in any other of the tested scenarios (Figure 3). Besides, some of the signals observed do not match any of the previously identified QTL regions and might have introduced errors in the genomic prediction equations. Similar patterns were observed for every trait showing decreased evaluation accuracy when sequence variants covering the whole chromosome were included.

In ssGBLUP, when the selection of variants is limited to the QTL region (between 24.72 and 28.38 Mb, 22,269 sequence variants selected), the gains in accuracy compared with the geno_50k scenario were high for traits associated with the region without a loss in accuracy for the other traits (TO and UP). The mean percentage of explained variance for the region between 23 and 30 Mb was 1.77% with the 50K genotypes alone but increased to 9.37% on average when sequence variants in the QTL region were added. In ssGBLUP, this increased percentage led to tremendous gains in accuracy especially for MY (13.7%), PY (12.5%), FY (17.9%), and UFP (9.0%) for which the variance explained by the region between 23 and 30 Mb reached 8.24, 9.03, 7.30, and 17.77%, respectively (Figure 2). However, we also observed in the geno_50kseqQTL scenario an increase in the percentage of variance explained by this region for traits that are not associated with this region (TO and UP): 4.1 and 4.6% for TO and UP, respectively. This increase might be an artifact linked to the enrichment of the area, each SNP having a small effect. This artifact could explain the deterioration of accuracy for these traits in this scenario. Besides, with the inclusion of sequence data, we introduce variants which are not completely independent leading to an over- or under-estimation of the variance they explain. In light of these findings, the updated chip appears as the best scenario as it has a lesser effect on the traits not associated with CHI19 than the sequence data of the QTL region (Figure 2) and as the choice of the markers partly took into account their LD.

Selecting variants within QTL regions therefore represents a great opportunity to improve routine genomic evaluations without losing in speed. Identifying the causal mutation for each trait might lead to further increases in the accuracy of genomic evaluations. However, pinpointing one variant is difficult given the number of variants in the area, the high LD and the low estimated recombination rates (lower than 1; R. Rupp, INRAE, personal communication). Nevertheless, Bolormaa et al. (2019) reported a significant effect of the accuracy of imputation to sequence level on the accuracy of genomic prediction in sheep data. It therefore

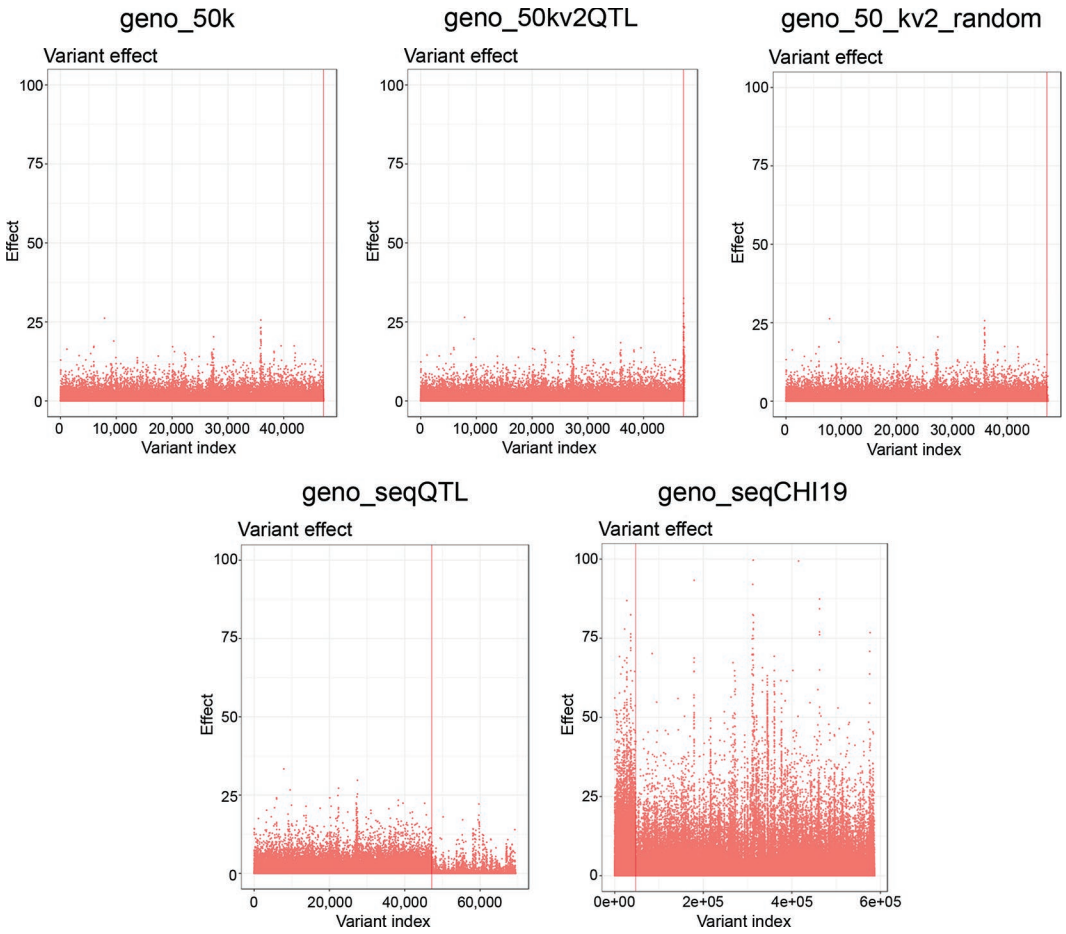

Figure 3. Variant effects in the single-step genomic BLUP scenarios tested for fore udder trait in the 146 validation Saanen individuals 50K markers on the left of the red line; additional variants on the right side of the red line.

seems likely that sequencing more individuals could further improve evaluations by lifting imputation errors and refining the QTL region.

Nevertheless, it is important to bear in mind that imputation has to be performed to get sequence information for every genotyped individual. This is a time-consuming process and imputed sequences represent a large amount of data that require significant storage capacities. In our study, we also show that the future version of the Illumina GoatSNP50 BeadChip is promising if the additional variants selected in the QTL region (178 SNP variants) are added for routine genomic evaluations. Indeed, the mean gain of accuracy was 4.9% with ssGBLUP and ranged from −0.1% (TO and LSCS; not significant) to 10.1% (FY). This updated chip has an increased density within the QTL region of chromosome 19 with 308 SNP (130 SNP from version 1 + 178 new selected variants) located between 23 and 30 Mb with an expected average spacing of 22,675 bp (ranging from 4 to 166,413 bp). However, the selected variants to be added to the chip have yet to be validated with a cluster file. A confirmation study will be needed with the real genotypes and the variants still present in the genotypes after the quality check before implementation in routine evaluation pipelines.

Similar studies were performed by VanRaden et al. (2017) on Holstein bulls. They achieved less significant gains in accuracy when sequence data were added to HD genotypes with only 0.6 percentage points when using solely SNP and 0.4 percentage points when adding both SNP and insertion/deletions (indels). This might be due to the fact that HD genotypes are already exhaustive compared with 50K genotypes so gains might be lower. In goats, the only genotyping tool available is a medium density chip, so it makes sense that the gains in accuracy when sequence data are added are higher than in dairy cattle. In the aforementioned study, adding 16,648 candidate SNP to the routinely used 60k SNP systematically resulted in an average gain that was higher for all traits. However, the average increase

was marginal and other studies also showed little gain when adding sequence information to 50K genotypes in cattle (B. O. Fragomeni et al., 2019).

Several studies have demonstrated that selecting a subset of variants to be included in the genomic evaluations is a particularly relevant approach. VanRaden et al. (2017) demonstrated that selecting a subset of SNP (almost 17,000 SNP) with the highest effects maximized the gains in accuracy (2.7 percentage points). Brøndum et al. (2015) selected 1,623 variants for the custom Illumina BovineLD SNP chip for Nordic breeds by performing association. They observed gains in accuracy ranging from 3 to 5 percentage points for production traits, less than 1 percentage point for udder health, and 0.5 percentage point for fertility. However, these gains are small compared with our findings. On simulated data, including QTL information seemed to increase the accuracy of the evaluations similarly to our results in the Saanen breed. Fragomeni et al. (2017) simulated livestock populations and obtained a gain of 8.2% when unweighted QTL information was added to a ssGBLUP model.

Model Comparisons

Our study follows the work of Teissier et al. (2019, 2019) who improved the accuracy of genomic evaluations of protein content by 4% when using WssGBLUP models taking the highest effect for a 40-SNP window. They also reported significant gains for milk yield (7 percentage points), fat and protein yields (4 and 5 percentage points, respectively), udder floor position (4 percentage points), rear udder attachment (1 percentage point) in Saanen goats compared with a ssGBLUP model. Our results show that using a standard WssGBLUP including sequence data from the QTL region slightly increased the accuracy of genomic evaluations by 2.1% on average compared with ssGBLUP performed on 50K genotypes. This is especially true for the production traits MY, FY, and PY for which the gains in accuracy were always higher than for other traits. This has also been reported in previous studies on 50K genotypes (Teissier et al., 2019). However, as previously observed by Teissier et al. (2018, 2019), the increase in gain is variable depending on the trait. Indeed, the accuracies for production traits are increased by 15% on average whereas other traits tended to show slight decreases. LSCS was the most affected trait with an evaluation accuracy that decreased by 15.8%. This might be related to the variant selection process for the update which selected variants significant for the highest number of traits but without checking the number of variants associated with each trait. Table 5 shows the representativeness of each trait with the variants selected for the update. LSCS is currently not represented on the updated version of the chip because of its low association to the QTL region. This might explain why the selected variants tend to introduce noise in the evaluations for this particular trait, and adding weights to the variants within the region might have amplified the phenomenon. Nevertheless, we reasonably question whether the ratio between the gain in accuracy and the time spent preparing the data (quality control, imputation, data formatting for software) is significant enough to implement the procedure on a routine basis.

Table 5. Representativeness of each trait associated with the QTL region of chromosome 19 (CHI19) among the 178 variants selected for the chip update

Trait	Number of variants
FU	12
UFP	150
RUA	52
LSCS	0
MY	104
FY	37
PY	62

[1]FU = fore udder; UFP = udder floor position; RUA = rear udder attachment; LSCS = 250-d SCS; MY = milk yield; FY = fat yield; PY = protein yield.

In the $WssGBLUP_{windows}$ models, regardless of the method used to build the windows, the QTL information tended to be smoothed (Figure 4). All weighted scenarios using sequence information from the QTL region (geno_50kseqQTL) led to significant decreases in the accuracy of genomic evaluations for both type traits (except for UFP and RUA) and LSCS compared with simple ssGBLUP on 50K genotypes. This finding might be caused by the use of extremely summarized information. Indeed, when windows were built on a distance basis (2.4 Mb), the whole chromosome 19 was spanned by only 26 windows and windows of 40 consecutive variants led to the construction of 586 windows on chromosome 19. These results are inconsistent with previous results in the French Saanen breed. Indeed, Teissier et al. (2018) improved genomic evaluations of the protein content when using windows of 40 consecutive variants. However, the casein region is smaller and has larger effects on the trait they studied in Saanen goats than chromosome 19, as shown by Carillier-Jacquin et al. (2016) who reported that the genetic variance explained by the αs1-casein gene reached 38% in this breed. The 50K SNP in the QTL region on chromosome 19 only explain between 6.6 and 21.5% of the genetic variance, depending on the trait. Its total effect might therefore be diluted when the region is divided into windows.

Regardless of the method used to build the windows, the variants selected for the chip update in the QTL

region led to an average gain in accuracy (6.4% for 2.4-Mb windows and 6.3% for 40-variant windows). The updated version of the chip therefore seems to perform slightly better than 50K genotypes (an increase of 2.2% on average) or the use of sequence information for the QTL region (up 4.2% on average). This result is promising; however, the gain is variable depending on the trait. Production traits are 17.7% more accurate when using $\text{WssGBLUP}_{\text{windows}}$. Other traits tended to show decreases in accuracy except for UFP, which increased 8%. Similarly to WssGBLUP, the traits that showed the highest gains with the 50K chip update in $\text{WssGBLUP}_{\text{windows}}$ models were those with the highest number of associated variants in the update (Table 5).

The aim of this study was to investigate whether sequence data could be included in routine genomic evaluations and if so, which model would provide optimal results. The findings we present here are preliminary and further investigation is needed to assess the influence of other known QTL regions in French dairy goats, for example the DGAT region (CHI14) for fat content and the casein region (CHI6) for protein content. Different models have already been studied in French dairy goats using solely 50K genotypes (Teissier et al., 2018, 2019) and in the light of our results the availability of sequence data represents a great perspective. When choosing the optimal strategy, one should bear in mind that production traits have a higher weight in the calculation of indexes than type traits (Larroque et al., 2011). Indeed, the formula for the Saanen breed is as follows:

Combined Index = Production Index + 0.6 × Morphology Index.An evaluation model that tends to improve the accuracy of prediction of production traits might therefore be preferable. It hence seems reasonable to exclude using ssGBLUP with the updated version of the chip as it significantly deteriorates the accuracy of evaluation of PY (down 7.4%). However, an ssGBLUP model using sequence data from the QTL seems appropriate. All WssGBLUP and $\text{WssGBLUP}_{\text{windows}}$ scenarios also increased the accuracy of prediction of production traits.

Even though type traits are of smaller importance in the combined index, scenarios and models that do not deteriorate the prediction of type traits should be favored. In the morphology index, each type trait is assigned the same weight. In this light, a WssGBLUP model with windows of 2.4 Mb on the updated version of the chip seems the best option as it is the model that causes the smallest decrease in the accuracy of prediction for LSCS, UP, and TO while improving the accuracy of genomic predictions for other traits.

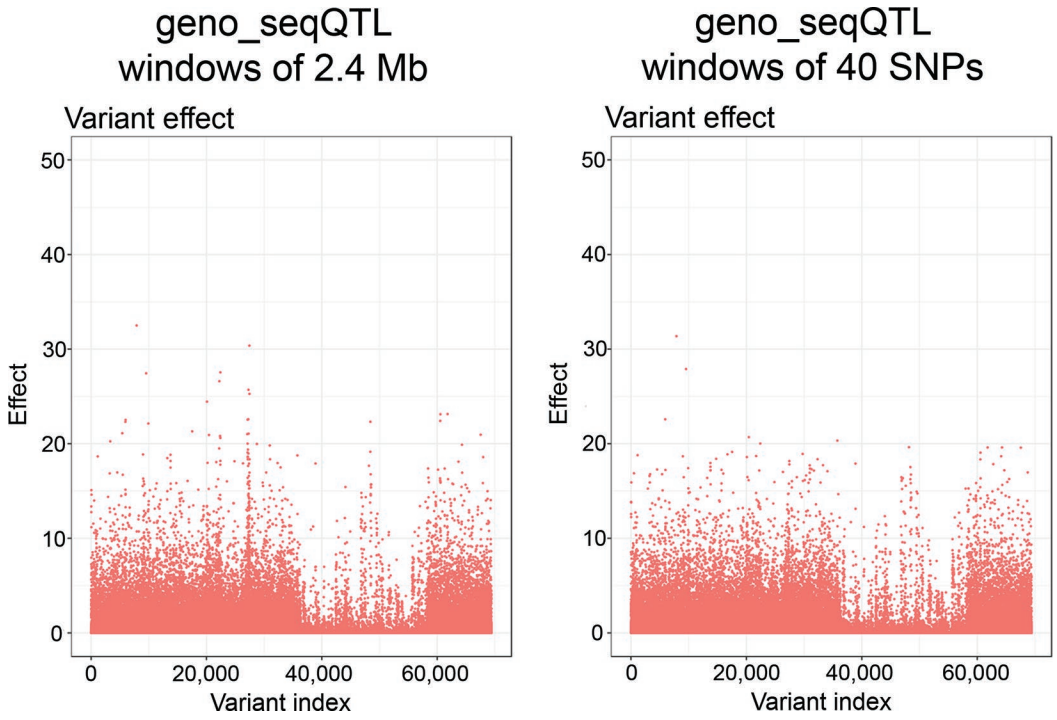

Figure 4. Variant effects in the weighted single-step genomic BLUP (WssGBLUP) scenarios using sequence variants of the QTL region for fore udder trait in the 146 validation Saanen individuals.

CONCLUSIONS

Including sequence data from the QTL region of chromosome 19 led to significant gains in accuracy for the genomic evaluation of traits associated with this region of the genome. The most time-efficient way to take sequence data into account on a routine basis seems to be a simple ssGBLUP model using variants of the QTL region. However, the upcoming chip update paves the way for a more strategic approach. Indeed, information from only a few carefully selected variants led to increased accuracies for the traits of interest. If production traits are to be emphasized, the WssGBLUP model makes the most of the updated chip with significant increases for MY, PY, and FY. Besides, this update is interesting as it would avoid time-consuming imputation and data formatting processes and provide reliable genotypes.

ACKNOWLEDGMENTS

This study would not have been possible without the sequence data provided by the VarGoats Consortium (http://www.goatgenome.org/vargoats.html) and previous work by the International Goat Genome Consortium (IGGC, http://www.goatgenome.org/) and ADAPTmap Consortium (http://www.goatadaptmap.org/) providing relevant DNA samples, genotyping tools, and genotyping data through their collaborative networks. Genotypes were funded by several projects: the French Genovicap and Phenofinlait programs (ANR, Apis-Gène, CASDAR, FranceAgriMer, France Génétique Elevage, the French Ministry of Agriculture Agrifood, and Forestry), the European 3SR project, and Maxi'male (CASDAR). We also thank the Capgenes breeding organization for the data provided. We are grateful to the Genotoul bioinformatics platform Toulouse MidiPyrénées and the CTIG (Centre de Traitement de l'Information Génétique) of INRAE Jouy-en-Josas, France, for providing computing resources. The authors thank Ignacy Misztal (University of Georgia, Athens, GA) for the blup90iod2 program. The first author also received financial support from the Occitanie region and the French Research National Research Institute for Agriculture, Food and Environment (INRAE – Animal Genetic division). CRG and RR designed the study. ET analyzed the data and drafted the manuscript. PB called the variants and provided support in computing. MT helped with the implementation of the different genomic models. HL, IP, and VC provided information on the current routine evaluations. IP provided part of the performance file and chose individuals to be sequenced. ET, GTK, CRG, and RR interpreted the results. RR and CRG improved the manuscript. All authors read and approved the final manuscript. The authors declare they do not have any competing interests.

REFERENCES

Bolormaa, S., A. J. Chamberlain, M. Khansefid, P. Stothard, A. A. Swan, B. Mason, C. P. Prowse-Wilkins, N. Duijvesteijn, N. Moghaddar, J. H. van der Werf, H. D. Daetwyler, and I. M. MacLeod. 2019. Accuracy of imputation to whole-genome sequence in sheep. Genet. Sel. Evol. 51:1. https://doi.org/10.1186/s12711-018-0443-5.

Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M. S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J. Dairy Sci. 98:4107–4116. https://doi.org/10.3168/jds.2014-9005.

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-Jacquin, C., H. Larroque, and C. Robert-Granié. 2016. Including α_{s1} casein gene information in genomic evaluations of French dairy goats. Genet. Sel. Evol. 48:54. https://doi.org/10.1186/s12711-016-0233-x.

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.

Clément, V., P. Martin, and F. Barillet. 2006. Elaboration of a total merit index combining dairy and udder type traits. Renc. Rech. Rumin. 1:209–212.

Fragomeni, B. O., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2017. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. Genet. Sel. Evol. 49:59. https://doi.org/10.1186/s12711-017-0335-0.

Fragomeni, B. O., D. A. L. Lourenco, A. Legarra, P. M. VanRaden, and I. Misztal. 2019. Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. J. Dairy Sci. 102:10012–10019. https://doi.org/10.3168/jds.2019-16262.

Hayes, B. J., I. M. Macleod, H. D. Daetwyler, P. J. Bowman, A. J. Chamberlian, C. J. Vander Jagt, A. Capitan, H. Pausch, P. Stothard, X. Liao, C. Schrooten, E. Mullaart, R. Fries, B. Guldbrandtsen, M. S. Lund, D. A. Boichard, R. F. Veerkamp, C. P. Vantassell, B. Gredler, T. Fruet, A. Bagnato, J. Vilkki, D. J. deKoning, E. Santus, and M. E. Goddard. 2014. Genomic Prediction from Whole Genome Sequence in Livestock: The 1000 Bull Genomes Project. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, 2014.

Hickey, J. M., B. P. Kinghorn, B. Tier, J. H. J. van der Werf, and M. A. Cleveland. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genet. Sel. Evol. 44:9. https://doi.org/10.1186/1297-9686-44-9.

Larroque, H., J. M. Astruc, A. Barbat, F. Barillet, D. Boichard, B. Bonaïti, V. Clément, 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), 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.

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. Tosser 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.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvrey, T. Druet, and D. Lee. 2002. BLUPF90 and related programs. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Moghaddar, N., I. M. Macleod, N. Duijvesteijn, S. Bolormaa, M. Khansefid, A. A. Swan, H. D. Daetwyler, and J. H. J. van der Werf. 2018. Genomic evaluation based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Proceedings of the World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.

Mucha, S., R. Mrode, M. Coffey, M. Kizilaslan, S. Desire, and J. Conington. 2018. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. J. Dairy Sci. 101:2213–2225. https://doi.org/10.3168/jds.2017-12919.

Pérez-Enciso, M., J. C. Rincón, and A. Legarra. 2015. Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised. Genet. Sel. Evol. 47:43. https://doi.org/10.1186/s12711-015-0117-5.

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.

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.

Talouarn, E., P. Bardou, I. Palhière, C. Oget, V. Clément, G. Tosser-Klopp, R. Rupp, and C. Robert-Granié. 2020. Genome wide association analysis on semen volume and milk yield using different strategies of imputation to whole genome sequence in French dairy goats. BMC Genet. 21:19. https://doi.org/10.1186/s12863-020-0826-9.

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-Granié. 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, W. Jun Wang, Wang, and W. Zhang. 2014. Design and characterization of a 52K SNP chip for goats. PLoS One 9:e86227. https://doi.org/10.1371/journal.pone.0086227.

VanRaden, P. M., M. E. Tooker, J. R. O'Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Genet. Sel. Evol. 49:32. https://doi.org/10.1186/s12711-017-0307-4.

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, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134. https://doi.org/10.3389/fgene.2014.00134.

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.

Williams, E. J. 1959. The comparison of regression variables. J. R. Stat. Soc. B 21:396–399. https://doi.org/10.1111/j.2517-6161.1959.tb00346.x.

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

Estelle Talouarn https://orcid.org/0000-0002-5016-0446
Marc Teissier https://orcid.org/0000-0002-0137-961X
Rachel Rupp https://orcid.org/0000-0003-3375-5816