A first step toward genomic selection in the multi-breed French dairy goat population

Céline Carillier, Helene H. Larroque, Isabelle Palhiere Palhière, Virginie Clément, Rachel Rupp, Christèle Robert-Granié

J. Dairy Sci. 96:1–12
http://dx.doi.org/10.3168/jds.2013-6789

A first step toward genomic selection in the multi-breed French dairy goat population

C. Carillier,*[1] H. Larroque,* I. Palhière,* V. Clément,† R. Rupp,* and C. Robert-Granié*
*Institut National de la Recherche Agronomique (INRA), UR631, Station d'Amélioration Génétique des Animaux (SAGA), 24 Chemin de Borde Rouge, Auzeville CS 52627, 31326 Castanet-Tolosan cedex, France
†Institut de l'élevage (Idele), 24 Chemin de Borde Rouge, Auzeville CS 52627, 31326 Castanet-Tolosan cedex, France

ABSTRACT

The objectives of this study were to describe, using the goat SNP50 BeadChip (Illumina Inc., San Diego, CA), molecular data for the French dairy goat population and compare the effect of using genomic information on breeding value accuracy in different reference populations. Several multi-breed (Alpine and Saanen) reference population sizes, including or excluding female genotypes (from 67 males to 677 males, and 1,985 females), were used. Genomic evaluations were performed using genomic best linear unbiased predictor for milk production traits, somatic cell score, and some udder type traits. At a marker distance of 50 kb, the average r^2 (squared correlation coefficient) value of linkage disequilibrium was 0.14, and persistence of linkage disequilibrium as correlation of r-values among Saanen and Alpine breeds was 0.56. Genomic evaluation accuracies obtained from cross validation ranged from 36 to 53%. Biases of these estimations assessed by regression coefficients (from 0.73 to 0.98) of phenotypes on genomic breeding values were higher for traits such as protein yield than for udder type traits. Using the reference population that included all males and females, accuracies of genomic breeding values derived from prediction error variances (model accuracy) obtained for young buck candidates without phenotypes ranged from 52 to 56%. This was lower than the average pedigree-derived breeding value accuracies obtained at birth for these males from the official genetic evaluation (62%). Adding females to the reference population of 677 males improved accuracy by 5 to 9% depending on the trait considered. Gains in model accuracies of genomic breeding values ranged from 1 to 7%, lower than reported in other studies. The gains in breeding value accuracy obtained using genomic information were not as good as expected because of the limited size (at most 677 males and 1,985 females) and the structure of the reference population.

Received March 11, 2013.
Accepted July 24, 2013.
[1]Corresponding author: celine.carillier@toulouse.inra.fr

Key words: dairy goat, genomic evaluation, linkage disequilibrium, female genotype

INTRODUCTION

Selection in the French Alpine and Saanen dairy goat breeds has been implemented by a single breeding organization. The objectives of this breeding scheme were to improve milk composition, milk yield, and udder morphology. Selection for these characteristics was based on a combined index calculated from EBV for milk yield, fat and protein yields, fat and protein contents, and various udder-type traits. This total merit index, which differs for Alpine and Saanen breeds (Clément et al., 2006), will change in 2013 to introduce selection on SCS (Rupp et al., 2011). Genetic evaluation of milk production traits has been carried out simultaneously in the 2 breeds using a BLUP animal model and considering all female performance records since 1980. Genetic evaluation of type traits (based on performance recorded since 2000) and SCS is performed separately for the 2 breeds.

The 2 breeds originated from the same single breed. The white coat variety of the Alpine goat, bred in the northern area of the Swiss Alps, was selected centuries ago to create the Saanen breed (Babo, 2000). When Alpine and Saanen were introduced in France in the 1910s, they were largely crossbred. In the 2000s, the percentage of Alpine genes in Saanen goats was 3.6 (Piacere et al., 2004), and genetic distance between the 2 breeds was <0.13 (Araujo et al., 2006).

The core selection population was composed of 1,000 dams of bucks selected each year for their reproductive ability, genetic level, and morphology. The AI rate was 20% in all goat herds. For health (e.g., no Q fever, tuberculosis), reproductive ability, growth, and genetic level, only 20% of all males (40 Alpine and 35 Saanen) born from assortative mating with the dams of bucks were progeny tested each year. Among those progeny-tested males, 25 Alpine and 15 Saanen were used as AI bucks. Progeny testing was performed on at least 60 daughters per buck over 18 mo. This led to a short generation interval of less than 4 yr in the sire-daughter

pathway, but a longer one (>5.5 yr) in the sire-son pathway (Danchin-Burge, 2011).

The availability of the Illumina goat SNP50 BeadChip (Illumina Inc., San Diego, CA; Tosser-Klopp et al., 2012) and recent genotyping methods means it is now feasible to assess genomic selection in this species. In dairy cattle, genomic selection has led to decreased generation intervals in the male pathway because of the early selection of males and improved breeding value accuracies for young animals at birth (Schaeffer, 2006; de Roos et al., 2010). Although the generation interval in the sire-son pathway in French dairy goats is shorter than that in dairy cattle, it is expected to be reduced with genomic selection, because of higher breeding value accuracies of young males at birth. In French dairy goats, AI bucks were largely used by breeders with more than 1,000 daughters per males, which led to accurate breeding values of males. Breeding values of young males at birth were 62% accurate on average, using average parent EBV accuracy. The aim of using genomic selection in this species would be to obtain breeding value accuracies for young males at least as accurate as the pedigree-derived breeding value accuracies to limit progeny testing.

The quality of genomic predictions depends on the number of phenotypes and genetic markers, heritability of traits, the reference population size (Goddard, 2009; Hayes and Goddard, 2010; Liu et al., 2011), relationships within the reference population, and relationships between reference and candidate populations (Habier et al., 2010).The number of bucks progeny tested each year in the French dairy goat breeding scheme limited the number of genotyped bucks for this study. The small reference population of this study consisted of all bucks progeny tested from 1993 to 2009. To maintain the link between the reference and candidate population, it was not possible to increase the number of male genotypes by genotyping more generations of young males. For this reason, we assessed the use of genotyped females on breeding value accuracy of candidates, using genotypes from commercial females available at the time of the study.

A first objective of this study was to examine the structure of the reference population by considering the level of linkage disequilibrium (**LD**) within the population and between the 2 breeds. A second objective was to study the size and structure of the reference population on the accuracy of genomic EBV (**GEBV**).

MATERIALS AND METHODS

Animals and Genotypes

The Saanen and Alpine purebred animals included in this study were obtained from 2 populations. The first population had previously been used for QTL detection for new traits such as fine milk composition (Maroteau et al., 2012). This population consisted mostly of females from commercial herds: 2,254 females (938 Saanen and 1,316 Alpine) born between 2008 and 2009 and their 20 sires (9 Saanen and 11 Alpine). The best goats of this population will be used in future goat breeding as dams of bucks. The second population genotyped was composed of 852 bucks (369 Saanen and 483 Alpine born between 1993 and 2011). All bucks were progeny tested except young buck candidates born between 2010 and 2011. From 2003, all French bucks progeny tested in the breeding scheme were genotyped, of which 60% were Alpine bucks.

Animals were genotyped using the Illumina goat SNP50 BeadChip (Tosser-Klopp et al., 2012). Of the 53,347 SNP on the chip, 46,959 were validated after quality control. Missing SNP genotypes were not imputed, and the marker effect of the missing SNP genotypes was set to zero for affected animals. Quality control consisted of obtaining a call rate threshold of 98%, a minor allele frequency >1%, and checking for Hardy-Weinberg equilibrium. Validation was carried out separately for the 2 breeds, and SNP that were not retained in both Saanen and Alpine breeds were discarded. Because of poor DNA quality and the animal call frequency threshold being set at 99%, 2,810 genotyped animals (1,164 Saanen and 1,646 Alpine) were available for this study.

Phenotypes

Five milk production traits were considered: milk yield, fat yield, and protein yield ($h^2 = 0.3$ for milk, fat, and protein yields) and fat content and protein content ($h^2 = 0.5$) (Bélichon et al., 1999). Somatic cell score ($h^2 = 0.20$; Rupp et al., 2011) and 5 udder type traits were also studied: udder floor position ($h^2 = 0.29$), udder shape ($h^2 = 0.32$), rear udder attachment ($h^2 = 0.27$), fore udder ($h^2 = 0.30$), and teat angle ($h^2 = 0.31$) (Clément et al., 2002). All heritabilities defined for SCS and udder type traits were averages from Alpine and Saanen breeds. In general, values for milk production traits of Saanen goats were different (Bélichon et al., 1999) from those of Alpine goats (783 vs. 733 kg for milk yield in Saanen and Alpine respectively; Institut de l'élevage, 2010).

The phenotypes used for the genomic evaluation were daughter yield deviations (**DYD**) for the 677 AI bucks and yield deviations (**YD**) for the 1,985 females. Yield deviations were calculated from the official genetic evaluation of January 2012 (Clément et al., 2002) using Genekit software (Ducrocq, 1998) as performance corrected for fixed effects. Daughter yield deviations were

calculated from DYD averages corrected for environmental effects and the merit of their dams (VanRaden et al., 2009). Each female of this study had 2 lactations; that is, 2 YD per female, weighted by 1 for first lactation and by 0.8 for second lactation, as in the official genetic evaluation. Each male's DYD was weighted by effective daughter contributions (**EDC**; Fikse and Banos, 2001), calculated from all daughters considered in the national genetic evaluation. The EDC were calculated separately in each breed for SCS and type traits, and simultaneously for the other traits using crEDC software (Sullivan, 2010). Average EDC ranged from 36.5 for the teat angle trait to 65.9 for milk yield in the whole phenotyped male population.

Cases Studied

The aim of this study was to examine the effect of population size and female genotypes on the accuracy of predictions, using several reference and candidate populations. The first population (A) consisted of 67 males in the reference population and 148 candidates. The candidates were young males born in 2010 and 2011 with no daughters at the time of the present study. The 67 reference males were all born between 1999 and 2009; 54% were Alpine and 46% were Saanen (Table 1; case A). Among the 148 candidates, 15 were half-sibs of genotyped females. This small reference population was used to investigate the usefulness of adding genotyped females in a small male population. The second population (B) consisted of a reference population with the same previous 67 males plus 1,985 females, and the same 148 candidates as in case A. Cases C and D consisted of the same animals as cases A and B, respectively, plus 610 males in the reference population and the same candidate population as in previous cases. All the additional males in cases C and D were related (ancestors or half sibs) to the males in cases A and B. Of the 610 additional males in the reference population, 26% were born before 2001 (Table 1). In this study, case B was not compared with case C, because it is difficult to compare the addition of both males and females because of the different accuracies of their phenotypes.

Description of the Population

Extent of LD and Persistence of LD Phases. Estimations of the extent of LD between markers in the whole reference population and in each breed, as well as estimations of the persistence of LD phases among Alpine and Saanen breeds were calculated. Because phases of chromosomes in this study were unknown, the extent of LD between markers was measured between genotypes for each pair of SNP within chromosome. Thus, the measure of LD used was the correlation across diploid genotypes as proposed by Rogers and Huff (2009):

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

where g_i and g_j were the genotypes at SNP i and j, respectively. Average r^2 (squared correlation coefficient) values were calculated for 20-kb intervals. For the persistence of LD phases between breeds, average cor-

Table 1. Number of genotyped males for each case (population) by year of birth

	Cases[1]			
	A and B		C and D	
Year	Alpine	Saanen	Alpine	Saanen
1993–1998	0	0	24	19
1999	0	2	23	21
2000	0	0	19	18
2001	0	0	23	25
2002	2	6	28	29
2003	7	3	36	28
2004	4	0	41	25
2005	7	5	42	29
2006	5	7	40	25
2007	3	3	40	29
2008	0	0	28	17
2009	8	5	40	28
2010	43	33	43	33
2011	43	29	43	29

[1]Cases A and B: 67 males in reference population, 148 male candidates; cases C and D: 677 males in reference population, 148 male candidates.

relations of signed r-values between Alpine and Saanen were derived over intervals of the same distance.

The interest in using r-values instead of r^2-values for persistence of phases between breeds was the use of a signed value. The r-value can be different in 2 breeds even if the absolute value is similar. The extent of LD and persistence of LD phases were evaluated for a population of 677 AI bucks and 148 young bucks (Table 1; cases C and D).

Relationships and Inbreeding Between and Within Populations. Inbreeding and relationship coefficients were calculated using Pedig software (Boichard, 2006) for both reference and candidate populations in all cases studied. The relationship coefficient between 2 animals is the probability that, at a given locus, the 2 individuals share alleles identical by descent from the same ancestor. The inbreeding coefficient of an individual is the probability that, at a given locus, an individual has received similar alleles from both parents. In this study, it was calculated using the Meuwissen and Luo (1992) method.

Statistical Model for Genomic Evaluation

To estimate GEBV for both females and males, genomic BLUP (**GBLUP**) using genomic BLUPf90 software (Misztal et al., 2002) was implemented. The mixed model considered was $\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$, where $\mathbf{y}$ is a vector of phenotypes weighted by EDC for males (DYD phenotypes) and the official weights of lactation (1 for first lactation and 0.8 for second lactation) for females (YD phenotypes), $\mathbf{X}$ is the incidence matrix relating breed effect ($\boldsymbol{\beta}$) to the individuals, $\mathbf{Z}$ is a design matrix allocating observations to breeding values ($\mathbf{u}$), and $\mathbf{e}$ is a vector of random normal errors. Genomic values $\mathbf{u}$ were normally distributed with $\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$, where $\mathbf{G}$ is the genomic relationship matrix as defined by Van Raden (2008):

$$\mathbf{G} = 0.95 \times \frac{\mathbf{WW}'}{2\sum_{j=1}^{p} q_j(1-q_j)} + 0.05 \times \mathbf{A},$$

where p is the number of loci considered, q_j is the frequency of an allele of marker j estimated across Alpine and Saanen, $\mathbf{W}$ is a centered incidence matrix of SNP genotypes, and $\mathbf{A}$ is the pedigree-based relationship matrix. The genomic relationship matrix was derived from genomic and pedigree relationships to make $\mathbf{G}$ and $\mathbf{A}$ compatible (Christensen et al., 2012). Combining pedigree and genomic information in the relationship matrix avoids bias in the hypothesis of no selection in the base generation, which is true considering the $\mathbf{A}$ matrix but not considering $\mathbf{G}$ (Legarra et al., 2009; Vitezica et al., 2011). The $\mathbf{G}$ matrix was computed using allele frequencies across breed for computation simplicity reasons. Although considering the difference of allele frequencies in breed reduced the relationship coefficient between distant individuals, it did not affect the results on accuracy of GEBV (Makgahlela et al., 2013). Single nucleotide polymorphism marker effects were assumed to have a prior normal distribution and mixed model equations were used with the genomic relationship matrix (VanRaden, 2008). To study the effect of including genotyped females, YD and DYD phenotypes were taken into account together in the model. By definition, variances of YD and DYD were not the same:

$$\mathrm{Var}(2DYD_i) = \sigma_u^2 + \frac{1}{d_i}\left(2\sigma_u^2 + 4\sigma_e^2\right)$$

and

$$\mathrm{Var}(YD) = \sigma_u^2 + \sigma_e^2,$$

where d_i is the EDC of animal i. To take into account this difference, each EDC is multiplied by a coefficient

$$k = \frac{\sigma_e^2}{2\sigma_u^2 + 4\sigma_e^2},$$

where σ_u^2 is the genetic variance and σ_e^2 is the residual variance.

Because of the small population size (i.e., <400 male genotypes available per breed), Alpine and Saanen populations were analyzed together, considering the 2 breeds as 1, as in Bélichon et al. (1999). The genetic parameters considered were the official parameters for milk production traits (Alpine and Saanen treated together) and the average parameters of Alpine and Saanen goats for SCS and type traits.

Accuracy and Bias of Genomic Evaluation. Accuracy and bias of genomic evaluation were estimated by splitting the 677 males from the total reference population into a training set and a validation set. The training set consisted of 425 males born between 1993 and 2005. The DYD of these males were obtained from official 2008 genetic evaluations. This set was used to predict the GEBV of the validation population of 252 males (i.e., 37% of total population) born between 2006 and 2009. Accuracies of genomic selection were derived from a correlation between GEBV and DYD of validation males, where DYD were estimated from official genetic evaluation of January 2012. Pedigree-derived accuracies of validation males were estimated from correlation between EBV and DYD. The EBV

were obtained from training males using the same model as for GEBV, except that $\text{Var}(u) = \mathbf{A}\sigma_u^2$, where **A** is the pedigree relationship matrix, and excluding genotype information. The prediction equation used to compute EBV of young males was derived from DYD (and YD) of reference animals and pedigree. The gain of using genomic information was derived from the difference between correlations between GEBV and DYD and correlations between EBV and DYD for validation males.

Taking into account that females were born in 2008 and 2009, and that no males with known phenotypes were born after them, the effect of adding females on genomic evaluation accuracy could not be estimated by cross validation. Model accuracies, derived from prediction error variance (**PEV**), were used to investigate the benefit of adding female genotypes.

Model Accuracy of Young Buck Breeding Values Estimated from PEV. Estimates of additive genetic values and PEV were obtained for all animals using the GBLUP model. The model accuracy considered (ρPEV) was

$$\rho_{PEV} = \sqrt{\frac{\sigma_u^2 - PEV}{\sigma_u^2}},$$

where PEV was the variance of prediction error and σ_u^2 is the genetic variance estimated on GEBV obtained (Bijma, 2012).

Average model accuracies were calculated for the 148 young buck candidates not yet progeny tested. Because the genomic relationship matrix was built using pedigree information, pedigree and genomic accuracy derived from PEV could be compared, calculating gain of model accuracy using genomic evaluation (Legarra et al., 2009; Christensen et al., 2012). The gain in model accuracy was estimated using genomic information for young bucks by comparing pedigree-derived model accuracy with genomic model accuracy. Pedigree-derived accuracy was obtained using the method described in the Accuracy and Bias of Genomic Evaluation section: using the same pedigree information and phenotypes of males and females from a reference population as for GBLUP evaluation but without molecular information. The objective of this study was to investigate how average genomic prediction accuracy varies with reference population size and addition of males or females.

RESULTS AND DISCUSSION

Description of the Reference Population

Extent of LD in the Population. Average r^2 calculated for each breed separately (Alpine, Saanen) and for the multi-breed population (Alpine + Saanen) as a function of marker distance are presented in Figure 1. For the 3 populations studied, average r^2 decreased with increasing marker distance. This decrease was less substantial for marker distances >150 kb. Average r^2 was constant for distances >1,200 kb, and was 0.07 and 0.04, respectively, for the 2 single breed populations and the multi-breed population. Extent of LD estimated in Saanen was close to that estimated in Alpine. Average r^2 values obtained in the multi-breed population were lower than in the single-breed populations. In this study, the extent of LD between 2 consecutive SNP (i.e., 50 kb: average distance between 2 SNP on the chip) was 0.17 for single-breed populations and 0.14 for the multi-breed population.

Average r^2 values in dairy goat at 50 kb were similar to the values reported in Lacaune sheep [0.12; G. Baloche, INRA-Station d'Amélioration Génétique des Animaux (SAGA), Toulouse, France, personal communication] but lower than those reported in Holstein dairy cattle (from 0.18 to 0.3, de Roos et al., 2008; Habier et al., 2010) and in Landrace, Duroc, Hampshire, and Yorkshire pigs (from 0.46 to 0.36, Badke et al., 2012). Similarities in LD extent estimated in Saanen and Alpine breeds could be explained by their common ancestor. The lower estimates of average r^2 in the multi-breed population (Alpine + Saanen) than in the single breed populations are in agreement with results in dairy cattle (Toosi et al., 2010; Hozé, 2012). In the European multi-breed dairy cattle population, extent of LD was 0.15 at 70 kb compared with 0.19 and 0.25 in Montbeliarde and Brown Swiss breeds, respectively (Hozé, 2012). As expected, the difference in LD extent between the multi-breed population and the single breed populations increased with marker distance. For small marker distances, it was due to the common origin of the 2 breeds. For higher marker distances, it could be associated with the management of Alpine and Saanen as purebred for more than 40 yr. Indeed, LD calculated for small marker distance, when fewer recombinations are possible, reflects the former history of breeds. For larger distances, extent of LD reflects more recent history (Hayes et al., 2003).

Using simulation, Habier et al. (2007) demonstrated that part of genomic accuracy was due to LD, using the decay of accuracy and LD per generation. In German Holstein dairy cattle (with LD extent of 0.3 for 60 kb), the part of accuracy due to LD ranged from 10% for protein yield with a reference population of 1,048 dairy bulls to 47% for fat yield with a reference population of 2,960 bulls (Habier et al., 2010). Based on these results, the relatively low extent of LD measured in the dairy goat population in the current study should not lead to high values of genomic evaluation accuracies. However,

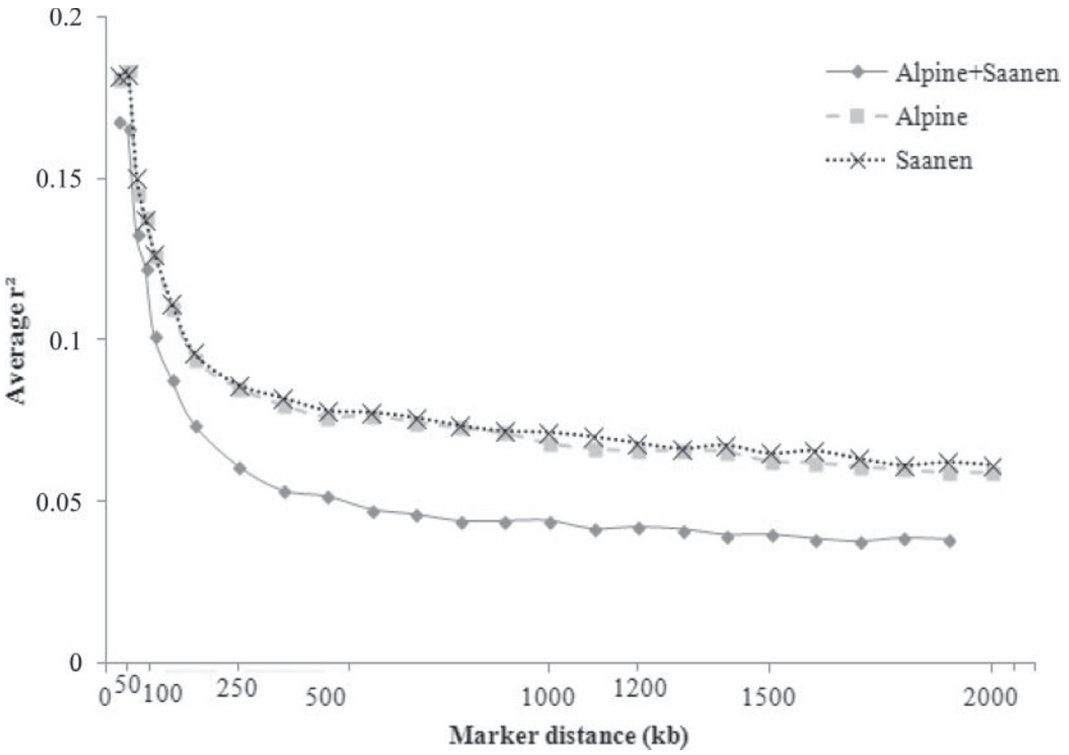

Figure 1. Linkage disequilibrium (average r^2) in Saanen and Alpine breeds and in the whole population (Alpine + Saanen).

accuracy of genomic evaluation was not the only parameter that influenced genomic evaluation accuracy.

Persistence of LD Phases Among Saanen and Alpine Breeds. Figure 2 shows the correlations between signed r-values of extent of LD among Alpine and Saanen breeds as a function of the distance between markers. Persistence of LD phases among Alpine and Saanen breeds decreased with genomic distance. At marker distance <50 kb, correlations of r among Alpine and Saanen breeds ranged from 0.88 to 0.56. This means that 2 SNP had the same level of LD in the Alpine breed and in the Saanen breed. The persistence of LD phases at 50 kb (i.e., average distance between 2 SNP) among Alpine and Saanen breeds was 0.56. Correlations of signed r-values estimated in Saanen and Alpine breeds decrease with increasing genomic distance between markers.

For short marker distances, persistence of LD phases among Alpine and Saanen was similar to the values reported between French Manech red-faced and black-faced sheep [0.5; G. Baloche, INRA-Station d'Amélioration Génétique des Animaux (SAGA), Toulouse, France, personal communication]. The 2 goat breeds (Alpine and Saanen) were genetically close centuries ago, as were the 2 Manech sheep breeds.

For greater marker distances, correlations of r-values in the reference population (0.08 for 600 kb) were close to those found between Lacaune and Manech black-faced sheep [0.09 for 600 kb; G. Baloche, INRA-Station d'Amélioration Génétique des Animaux (SAGA), Toulouse, France, personal communication]. But they were lower than that reported in dairy cattle between Jersey and Holstein for 600 kb (de Roos et al., 2008), in beef cattle between Charolais and Angus (Lu et al., 2012), and between Landrace and Yorkshire pigs (Badke et al., 2012).

Combining several breeds in a single reference population was considered when persistence of LD phases was high, as for Jersey and Holstein. However, the moderate level of LD phase persistence for 2 consecutive markers in Alpine and Saanen goats did not prohibit combining both breeds.

Relationships and Inbreeding Between and Within Populations. The pedigree file of 37,669 animals common to the 2 breeds took into account 26 generations. The kinship coefficient in the whole population was, on average, 1.6%. The kinship coefficient calculated within the 4 reference populations ranged from 1.6 to 1.8% (Table 2). The highest coefficients were obtained for case B, because of the addition of daughters of males from case A, and for case C, because of the addition of strongly related males from case A. Nevertheless, the addition of the daughters of 20 males from case C to case D did not increase the kinship

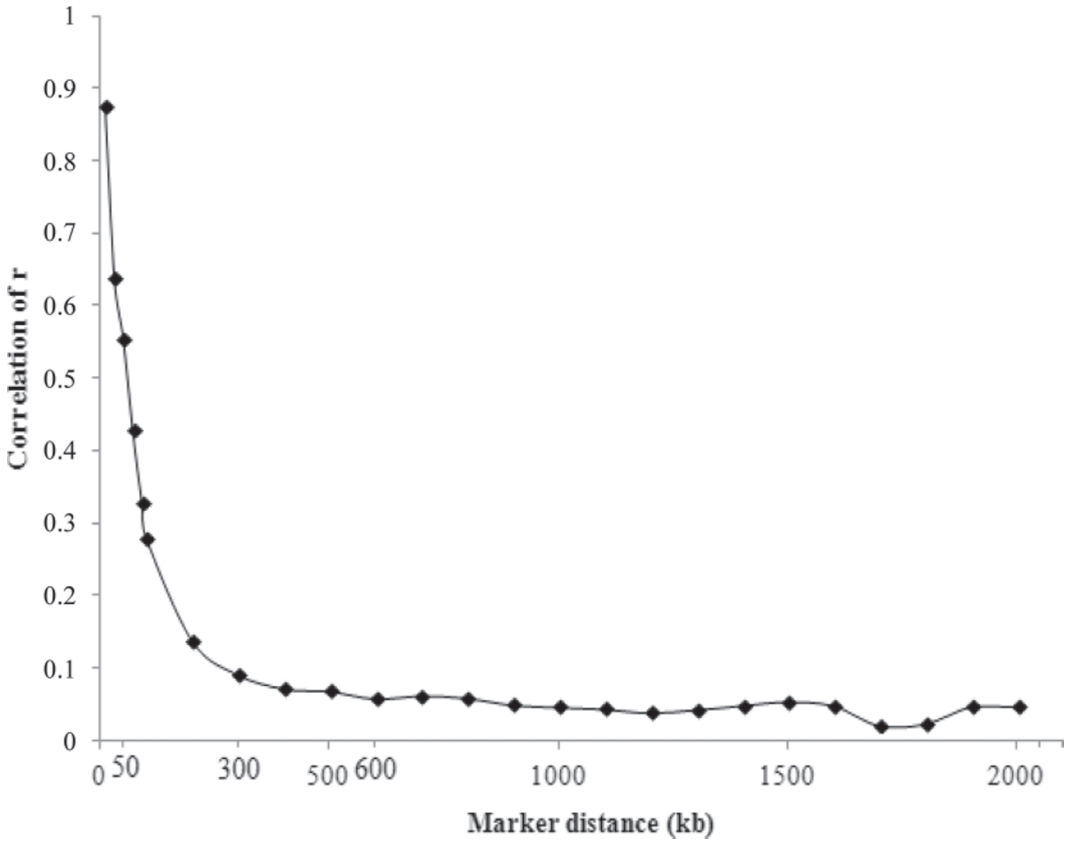

Figure 2. Persistence of linkage disequilibrium (LD) among Alpine and Saanen breeds derived as correlation of signed r LD values between the 2 breeds.

coefficient, because these females were not strongly related to all of the males. Kinship coefficients were 1.3% within the candidates and ranged from 1.3 to 1.4% between reference and candidate populations. Kinship coefficients reported in this study were low compared with that observed in cows (~5%; Habier et al., 2010; Pszczola et al., 2012).

Inbreeding coefficients (F_z in Table 2) were 2.1, 2.8, 2.2, and 2.6% within reference populations for cases A, B, C, and D, respectively, and 2.1% within candidate populations. They were lower than in Holstein dairy cattle (Miglior, 2000). The reported inbreeding in case B was caused by a higher proportion of females in the population, these being more inbred than the males. The relatively low levels of kinship and inbreeding coefficients within the populations can be attributed to the implementation of a new scheme of inbreeding management (optimizing contribution methods) in the French goat population in 2002 (Colleau et al., 2004). In this scheme, selection of AI bucks is managed within families to maintain genetic progress and minimize the average pairwise relationship coefficient in the population. However, among the 20 sires of the females, only a few had different ancestral origins. Inbreeding coefficients of those males and their daughters were higher (2.8 compared with 1.9%; results not shown) than those observed in the other males of the study.

Genomic Evaluation

Accuracy, Bias, and Gain in Accuracy of Genomic Evaluation Estimated by Cross Validation. Correlations between DYD and GEBV estimated in the validation population of 252 males ranged from 32.1% for SCS to 53.3% for fat content (Table3). The highest correlations were obtained for the most heritable traits (i.e., fat and protein contents).

These results were lower than those reported in the French Holstein dairy cattle population (Fritz et al., 2010) for similar traits: 39 versus 59% for milk yield, 36 versus 60% for protein yield, and 37 versus 63%, for udder floor position. These differences in accuracy could not be explained by DYD accuracy in French dairy goats (average EDC of 390), which is slightly higher than in dairy cattle, but are explained by the structure and the size of the reference population. Correlations

Table 2. Average (μ) and standard error (SE) kinship (F_{ij}) and inbreeding (F_z) coefficients within reference and candidate populations and between candidate and reference populations

	Case[1]							
	A		B		C		D	
Item	F_{ij}	F_z	F_{ij}	F_z	F_{ij}	F_z	F_{ij}	F_z
Within reference population								
μ (%)	1.6	2.1	1.8	2.8	1.8	2.2	1.7	2.6
SE (%)	2.1	1.0	3.2	1.4	2.6	1.1	2.9	1.3
Within candidate population								
μ (%)	1.3	2.1	1.3	2.1	1.3	2.1	1.3	2.1
SE (%)	2.3	0.7	2.3	0.7	2.3	0.7	2.3	0.7
Between reference and candidate populations								
μ (%)	1.4	—	1.3	—	1.4	—	1.3	—
SE (%)	2.1	—	2.0	—	2.2	—	2.0	—

[1]Case A: 67 males in the reference population, 148 male candidates; Case B: 67 males and 1,985 females in the reference population, 148 male candidates; Case C: 677 males in the reference population, 148 male candidates; Case D: 677 males and 1,985 females in the reference population, 148 male candidates.

between DYD and GEBV for our validation bucks were similar to those reported in Manech red-faced dairy rams (38 and 37% for milk and fat yields) with a training population of around 1,000 rams (Barillet et al., 2012) and in Normande dairy bulls (36 and 33% for milk and protein yields) with 930 training bulls (Fritz et al., 2010). Nevertheless, the results for SCS and type traits, between 32 and 43%, were slightly lower than in other species (48% for SCS of Manech, Barillet et al., 2012; 47% for udder floor position of Normande breed, Fritz et al., 2010).

Gains of accuracy using genomic information in our study (Table 3) ranged from 3.4% for protein content to 21.3% for fore udder. These gains for milk production traits were lower than those obtained for milk yield in other species (41% in Manech, Barillet et al., 2012; 41.7% in Normande, Fritz et al., 2010). This finding can be explained by the high pedigree-derived accuracies of young buck breeding values for those traits because of a high number of daughters per sire (388 in average). For udder floor position, gains were similar to those reported in the Normande dairy cattle breed (23.7%) from Fritz et al. (2012).

Regression coefficients presented in Table 3 ranged from 0.73 to 0.96. They were higher for fat and protein contents (96.2 and 94.9%), close to those reported in French Lacaune dairy sheep (85 to 86%; Duchemin et al., 2012) and French dairy cattle (71 to 113%; Karoui et al., 2012). A coefficient of 1 (indicating the absence of bias) was expected if the animals in the validation set were not selected. Biases of genomic breeding value estimations were low for fat and protein contents and for type traits, with regression coefficients up to 90%, except for fore udder.

Model Accuracy of Genomic Predictions for Candidates Estimated from PEV. Figure 3 shows the average model accuracy of genomic prediction (derived from PEV) calculated for the 148 candidates without progeny test results, in each case for milk yield (identical results were obtained for protein and fat yields), fat content (identical results were obtained for protein content), and teat angle (identical results were

Table 3. Correlations between daughter yield deviations (DYD) and genomic (G)EBV for males from validation population and regression coefficient of DYD onto GEBV

Trait	DYD × GEBV	DYD × EBV	Gain (%)	Regression coefficient
Milk yield	0.391	0.372	5.1	0.786
Fat yield	0.373	0.350	6.2	0.784
Protein yield	0.362	0.345	4.9	0.762
Fat content	0.533	0.495	7.7	0.962
Protein content	0.519	0.502	3.4	0.949
Somatic cell score	0.321	0.305	5.2	0.742
Udder floor position	0.367	0.304	20.7	0.918
Udder shape	0.339	0.280	21.1	0.899
Rear udder attachment	0.425	0.396	7.3	0.923
Fore udder	0.325	0.268	21.3	0.726
Teat angle	0.352	0.324	8.6	0.908

obtained for all type traits and SCS). Similar results were obtained when heritability and DYD accuracy were the same. For SCS and type traits, accuracy of DYD (EDC) for all of those traits was 35% lower than EDC of milk production traits. Accuracies ranged from 9% for type traits and SCS in case A to 56% for fat and protein contents in case D with highest heritabilities. The lower values observed for GEBV accuracies for the young males in case A could be explained by the small size of the reference population and the absence of all fathers of candidates in the reference population.

The highest accuracies obtained with a reference population of 677 males and 1,985 females were similar to those observed in Merinos sheep (ρPEV from 50 to 57%, for ultrasound-scanned traits) with an average relationship of 0.5 between animals (Clark et al., 2012) and in Jersey dairy cattle (from 52 to 57% for milk production traits; Hayes et al., 2009). These results were higher than those reported in hens (ρPEV: 42%, for presence test of *Salmonella* in spleen) because of the small number (1,342) of SNP used in Calenge et al. (2011).

The model accuracies obtained for the 257 males of the training population used in the previous section were lower than the accuracies derived from cross validation (from 9 to 56%, results not shown) as in other studies (Clark et al., 2012). This could be explained by not taking into account the genetic selection of candidates, which led to overestimations of model accuracy (Gorjanc et al., 2012).

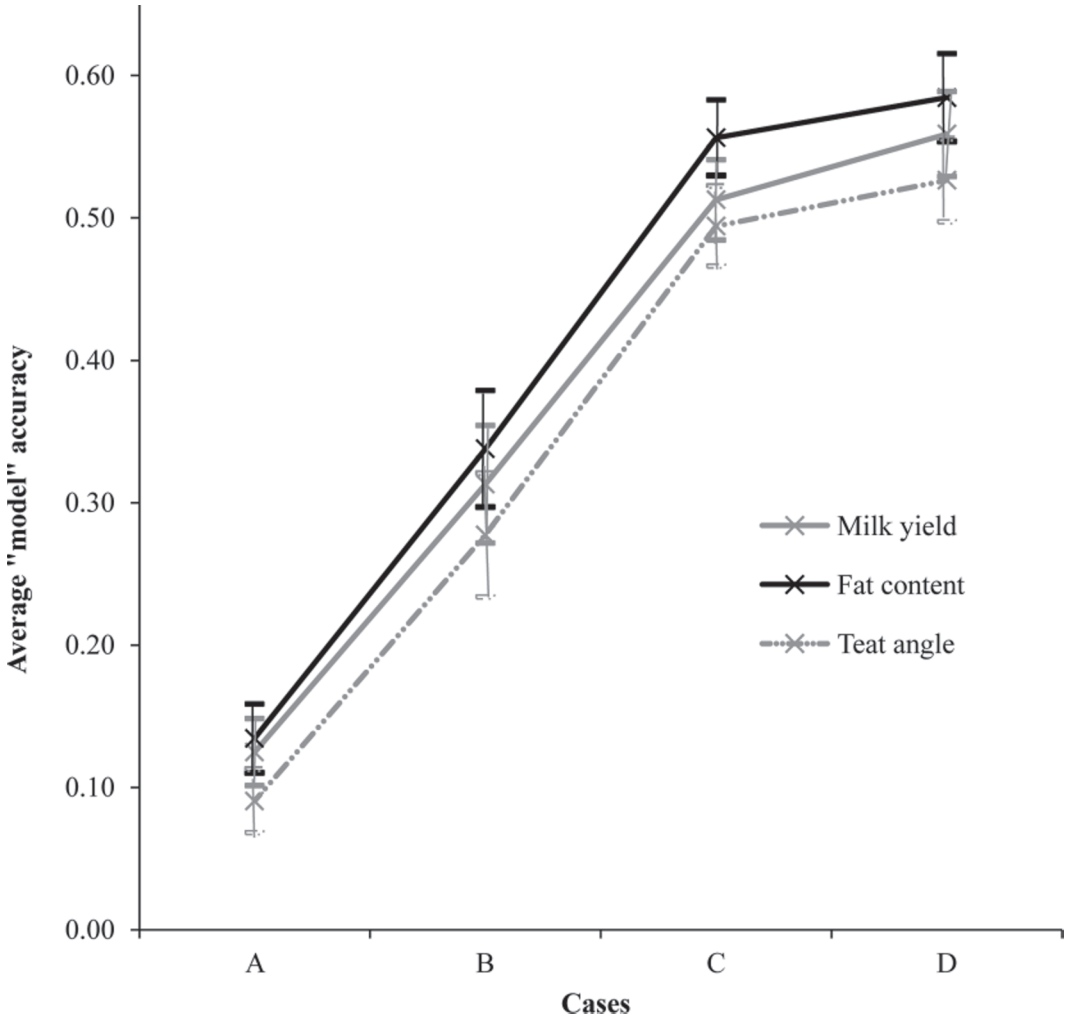

Figure 3. Average "model" accuracy derived from prediction error variance of genomic predictions for candidates in fat content, milk yield, SCS, and rear udder attachment. Case A: 67 males in the reference population, 148 male candidates; case B: 67 males and 1,985 females in the reference population, 148 male candidates; case C: 677 males in the reference population, 148 male candidates; case D: 677 males and 1,985 females in the reference population, 148 male candidates. The same results were obtained for protein yield and fat yield as are shown for milk yield; the same results were obtained for protein content as are shown for fat content; and the same results were obtained for udder floor position, udder shape, and fore udder as are shown for teat angle.

Table 4. Average (μ) and standard error (SE) of differences between breeding value "model" accuracies and pedigree accuracies derived from prediction error variance for the 148 male candidates

	Case[1]			
Trait	A	B	C	D
Milk yield[2]				
μ	0.03	0.05	0.03	0.05
SE	0.01	0.03	0.02	0.03
Fat content[3]				
μ	0.05	0.04	0.03	0.07
SE	0.02	0.01	0.02	0.04
Somatic cell score[4]				
μ	0.01	0.06	0.03	0.04
SE	0.02	0.02	0.03	0.03

[1]Case A: 67 males in the reference population, 148 male candidates; Case B: 67 males and 1,985 females in the reference population, 148 male candidates; Case C: 677 males in the reference population, 148 male candidates; Case D: 677 males and 1,985 females in the reference population, 148 male candidates.

[2]The same results were obtained for protein yield and fat yield.

[3]The same results were obtained for protein content.

[4]The same results were obtained for udder floor position, udder shape, fore udder, and teat angle.

Adding animals increased GEBV accuracy for all traits and in all cases (Figure 3). The addition of females increased the accuracy by 5% (case C vs. D). But accuracies for udder type traits and SCS increased greatly by 206% (case A vs. B, Figure 3) when females were added to the population of 67 males. However, in case D, the addition of females led to a greater improvement of GEBV accuracies for their 15 half-sibs than for the 148 other candidates; that is +30% for SCS (results not shown). In dairy cattle, the addition of dams of bulls slightly altered GEBV reliabilities derived from correlation between DYD and GEBV for candidate bulls, from −4.9% for fat yield in Holstein (Dassonneville et al., 2013) to 5.8% (Wiggans et al., 2011) and 8% (Pryce et al., 2012) for protein and fat contents. These lesser values could be due to the preferential treatment of some cows and lead to errors in the phenotypes of the bulls (Dassonneville et al., 2013); this is not the case in large herds such as in goats.

Gain in "Model" Accuracy of Genomic Predictions for Candidates. Table 4 shows the gains in the accuracy of EBV of the 148 candidates observed when using genomic information or only pedigree and phenotypic information. These gains ranged from 1% for SCS and type traits in case A to 7% for fat and protein contents in case D. These values were similar to gains of theoretical accuracy obtained for the presence of *Salmonella* in hens (from 0 to 15%; Calenge et al., 2011). The gains observed in this study were lower than those obtained for Merinos sheep: from 76% for ultrasound-scanned eye muscle depth to 468% for ultrasound-scanned traits (Clark et al., 2012).

The addition of females to the reference population of 67 males was less advantageous for gains in GEBV accuracy (+33% for SCS and type traits, case B vs. case A; Table 4) than adding them in a larger reference population (+500% for SCS and type traits, case D vs. case C; Table 4). This improvement in accuracy gains could be explained by an increase in GEBV accuracy between case C and case D, whereas EBV accuracy (calculated using only phenotypes and pedigree information) was similar in both cases.

All females used in this study were sired by 20 sires. Genetic diversity of these females was not wide enough that the addition of genotyped females actually improved the prediction model. An interesting point for the future would be to examine the benefit of genotyping a set of buck dams chosen to represent genetic diversity of the whole set of dams.

The current reference population of 677 males used in this study comprised all bucks progeny tested in the breeding scheme. Adding new generations of genotyped males to the reference population will increase the reference population size but will not improve relationship between the candidate individuals and the reference individuals. The main way to increase size and improve structure of the French dairy goat reference population could be done essentially by genotyping females. The choice of these females should be further investigated.

CONCLUSIONS

The present study is the first report to be published on genomic evaluation in dairy goats. The results describe the characterization of the French reference population available currently with an extent of LD of 0.14 between 2 consecutive SNP. Accuracies of genomic evaluation were similar to values reported in other

species, but gains in using genomic information were slightly low because of the structure and size of the reference population. Accuracies and gains in accuracy could be improved by adding genotyped females. The use of the multiple-trait model, models using haplotypes instead of SNP, and single-step genomic BLUP models will be examined in the future.

ACKNOWLEDGMENTS

This work was funded by the French Genovicap and Phenofinlait programs (ANR, Apis-Gène, CASDAR, FranceAgriMer, France Génétique Élevage, and French Ministry of Agriculture, Paris, France) and the European 3SR project. The first author benefitted from financial support from the Midi-Pyrénées region (Toulouse, France) and the SELGEN program of the French National Institute of Research in Agronomy (INRA, Paris, France). We thank Helen Munduteguy (Pyratus, Tarbes, France) and Wendy Brand-Williams (INRA, Jouy-en-Josas, France) for checking the English of our article and the two reviewers for their very constructive comments. This study would not have been possible without the goat SNP50 BeadChip developed by the International Goat Genome Consortium (IGGC; www.goatgenome.org).

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Araujo, A. M. D., S. E. F. Guimaraess, T. M. M. Machado, P. S. V. Lopes, C. S. Pereira, F. L. R. D. Silva, M. T. Rodrigues, V. D. S. Columbiano, and C. G. A. D. Fonseca. 2006. Genetic diversity between herds of Alpine and Saanen dairy goats and the naturalized Brazilian Moxoto breed. Genet. Mol. Biol. 29:67–74.

Babo, D. 2000. Les Races Ovines et Caprines Françaises. 1st ed. France Agricole, Paris, France.

Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel. 2012. Estimation of linkage disequilibrium in four US pig breeds. BMC Genomics 13:24.

Barillet, F., G. Baloche, G. Lagriffoul, H. Larroque, R. Robert-Granié, A. Legarra, and J. M. Astruc. 2012. Genomic selection in French Lacaune and Manech dairy sheep breeds: comparison of BLUP and GBLUP accuracies. Page 3 in Proc. 38th ICAR session, Cork, Ireland.

Bélichon, S., E. Manfredi, and A. Piacère. 1999. Genetic parameters of dairy traits in the Alpine and Saanen goat breeds. Genet. Sel. Evol. 31:529–534.

Boichard, D. 2006. Pedig, logiciel d'analyse de généalogies adapté à de grandes populations; version 2007. INRA SGQA, Jouy-en-Josas, France.

Calenge, F., A. Legarra, and C. Beaumont. 2011. Genomic selection for carrier-state resistance in chicken commercial lines. BMC Proc. 5(Suppl. 4):S24.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571.

Clark, S. A., J. M. Hickey, H. D. Daetwyler, and J. H. J. Van Der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet. Sel. Evol. 44:4. http://dx.doi.org/10.1186/1297-9686-44-4.

Clément, V., D. Boichard, A. Piacere, A. Barbat, and E. Manfredi. 2002. Genetic evaluation of French goats for dairy and type traits. Pages 235–238 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

Clément, V., P. Martin, and F. Barillet. 2006. Elaboration d'un index synthétique caprin combinant les caractères laitiers et des caractères de morphologie mammaire. Renc. Rech. Rumin. 13:209–212.

Colleau, J. J., S. Moureaux, M. Briend, and J. Béchu. 2004. A method for the dynamic management of genetic variability in dairy cattle. Genet. Sel. Evol. 36:373–394.

Danchin-Burge, C. 2011. Bilan de variabilité génétique de 9 races de petits ruminants laitiers et à toison. Compte rendu no. 001172004, Institut de l'Élevage collection résultats, juin 2011. Institut de l'Élevage, Paris, France.

Dassonneville, R., A. Baur, S. Fritz, D. Boichard, and V. Ducrocq. 2013. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. Genet. Sel. Evol. 45:40. http://dx.doi.org/10.1186/1297-9686-44-40.

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

de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2010. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94:1559–1567.

Duchemin, S. I., C. Colombani, A. Legarra, G. Baloche, H. Larroque, J.-M. Astruc, F. Barillet, C. Robert-Granié, and E. Manfredi. 2012. Genomic selection in the French Lacaune dairy sheep breed. J. Dairy Sci. 95:2723–2733.

Ducrocq, V. 1998. Genekit, BLUP software; June 2011 version. INRA SGQA, Jouy-en-Josas, France.

Fikse, W. F., and G. Banos. 2001. Weighting factors of sire daughter information in international genetic evaluations. J. Dairy Sci. 84:1759–1767.

FritzS.GuillaumeF.CroiseauP.BaurA.HozeC.DassonnevilleR.BoscherM. Y.JourneauxL.BoichardD. DucrocqV. 2010. Mise en place de la sélection génomique dans les trois principales races françaises de bovins laitier. Renc. Rech. Rumin. 17:455–458.

Goddard, M. E. 2009. Genomic selection: Prediction of accuracy and maximisation of long-term response. Genetica 136:245–257.

Gorjanc, G., J. M. Hickey, and P. Bijma. 2012. Reliability of breeding values in selected populations. Page 15 in Interbull Bulletin, 5, Cork, Ireland.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5. http://dx.doi.org/10.1186/1297-9686-42-5.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and E. Goddard. 2009. Accuracy of genomic breeding values in multibreed dairy cattle populations. Genet. Sel. Evol. 41:51. http://dx.doi.org/10.1186/1297-9686-41-51.

Hayes, B. J., and M. E. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. Genome 53:876–883.

Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13:635–643.

Hozé, C. 2012. High density chip brings new opportunity for multibreed genomic evaluations in dairy cattle. Page 6 in 16th QTL MAS Workshop, Alghero, Sardinia, Italy.

Institut de l'élevage. 2010. Résultats de Contrôle Laitier des espèces bovine et caprine (campagne 2010). Accessed Mar. 27, 2013. http://idele.fr/recherche/publication/idelesolr/recommends/resultats-de-controle-laitier-des-especes-bovine-et-caprine-campagne-2010.html.

Karoui, S., M. J. Carabano, C. Diaz, and A. Legarra. 2012. Joint evaluation of French dairy cattle breeds using multiple-trait models. Genet. Sel. Evol. 44:39. http://dx.doi.org/10.1186/1297-9686-44-39.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., and V. Ducrocq. 2012. Computational strategies for national integration of phenotypic, genomic and pedigree data in a single-step best linear unbiased prediction. J. Dairy Sci. 95:4629–4645.

Liu, Z., F. R. Seefried, F. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43:19. http://dx.doi.org/10.1186/1297-9686-43-19.

Lu, D., M. Sargolzaei, M. Kelly, C. Li, G. Vander Voort, Z. Wang, G. Plastow, S. Moore, and S. P. Miller. 2012. Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. Front. Genet. 3:1–10.

Makgahlela, M. L., I. Strandén, U. S. Nielsen, M. J. Sillanpää, and E. A. Mäntysaari. 2013. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. J. Dairy Sci. 96:5364–5375. http://dx.doi.org/10.3168/jds.2012-6523.

Maroteau, C., I. Palhière, H. Larroque, V. Clément, G. Tosser-Klopp, and R. Rupp. 2012. Genetic parameter estimation for major fatty acids in French dairy goats. Page 366 in 63rd Annu. Mtg. European Federation of Animal Science (EAAP), Bratislava, Slovakia.

Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305–313.

Miglior, F. 2000. Impact of inbreeding-Managing a declining Holstein gene pool. Pages 108–113 in Proc. 10th World Holstein Friesian Federation Conf., Sydney, Australia.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 743–745 in Session 28, Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

Piacere, A., I. Palhière, H. Rochambeau, and D. Allain. 2004. Analysis of the genetic variability of the French Alpine and Saanen breeds using genealogical data. Page 30 in Proc. 8th Int. Goat Conf., Pretoria, South Africa.

Pryce, J., B. Hayes, and M. E. Goddard. 2012. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. Page 28 in Proc. 38th ICAR Conf., Cork, Ireland.

Pszczola, M., T. Strabel, H. A. Mulder, and M. P. L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. J. Dairy Sci. 95:389–400.

Rogers, A. R., and C. Huff. 2009. Linkage disequilibrium between loci with unknown phase. Genetics 182:839–844.

Rupp, R., V. Clément, A. Piacere, C. Robert-Granié, and E. Manfredi. 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. J. Dairy Sci. 94:3629–3634.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223.

Solberg, T. R., A. K. Sonesson, and J. A. Woolliams. 2008. Genomic selection using different marker types and densities. J. Anim. Sci. 86:2447–2454.

Sullivan, P. 2010. Description of Usage for CrEDC_5e.c. Canadian Dairy Network, Guelph, Canada.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32–46.

Tosser-Klopp, G., P. Bardou, C. Cabau, A. Eggen, T. Faraut, H. Heuven, S. Jamli, C. Klopp, C. T. Lawley, J. Mcewan, P. Martin, C. Moreno, P. Mulsant, I. Nabihoudine, E. Pailhoux, I. Palhière, R. Rupp, J. Sarry, B. Sayre, A. Tircazes, J. Wang, W. Wang, T. P. Yu, and W. Zhang. 2012. Goat genome assembly, Availability of an international 50K SNP chip and RH panel: An update of the International Goat Genome Consortium projects. Pages 1–14 in Plant and Animal Genome Conf., San Diego, CA.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M., C. P. Van Tassel, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb.) 93:357–366.

Wiggans, G. R., T. A. Cooper, P. M. Vanraden, and J. B. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94:6188–6193.