European Master in Animal Breeding and Genetics (EM-ABG)
Master Sciences et Technologies du Vivant et de l’Environnement
Spécialité : Génétique Animale, Génome et Diversité

Evaluation of a reference population in dairy goats for genomic selection

Céline CARILLIER

Supervisors:
Virginie Clément, Hélène Larroque, Isabelle Palhière, Christèle Robert-Granié, Rachel Rupp

June 2012
Abstract

Reference population is the main point to succeed in genomic selection, improving genetic breeding value accuracies. The aim of this study was to characterize the structure of the multi-breed (Alpine and Saanen) dairy goat reference population genotyped with a 50k chip, studying level of relationship and linkage disequilibrium extent within population. Several reference populations (from 20 males to 705 males and 1 946 females) were designed to examine effect of population size and interest of female genotypes on genomic estimated breeding values (GEBV) using GBLUP. Feasibility of genomic selection in goats was considered comparing GEBV accuracies and ascendance accuracy calculated for young males. The reference population considered in this study with 250 effective founders and 0.115 level of LD extent between two consecutive markers was not perfect for genomic selection. Increase population size induced proportional increase in GEBV accuracies from +2% to +31% according to the case studied and the trait. Adding female genotypes improved accuracy from +0.28 to +0.01 depending on the reference population and trait studied, more relevant considering small population size. Genomic selection did not improve GEBV accuracies for young males except for those with sires genotyped.

Key words: Dairy goats, multi-bred reference population, female genotypes, linkage disequilibrium, genomic selection accuracies
Abstract

Genomic selection is done in several species like in bovine since 2010. In dairy goat, a 50k chip and genotypes are recently available. Interest of genomic selection lies in reduction of generation interval, already quite small in goats, and in improving accuracy of genetic breeding value of animals especially at birth. In this study, several multi-breed (Alpine and Saanen) reference populations, considering several population sizes with or without female genotypes (from 20 males to 705 males and 1,946 females), were used to compare accuracies of genomic estimated breeding values (GEBV). A characterization of the entire reference population structure, estimating level of relationship, LD extent and persistence between breeds was used to estimate GEBV accuracy level expected. Then genomic evaluation using GBLUP were done on several populations considered. The whole reference population studied was 705 males and 1,946 females. The traits studied were milk production traits, somatic cell score (SCS) and type traits. Effect of population size and interest of female genotypes were examined on the genomic accuracies. Feasibility of genomic selection was studied by comparing GEBV accuracies to ascendance accuracies calculated for young males. The high number of founders, 250, and the low LD extent between two consecutive markers, 0.115, in this reference population was not perfect for genomic selection. But results on persistence of LD phases, the same as in Manech red faced and black faced for low marker distance, justified using Alpine and Saanen in a unique reference population.

The population size had an important effect on GEBV accuracies, from +2% to +31% according to the case studied and the trait. Increasing it with females would be more interesting considering small population size, +0.15 with 93 animals, than bigger one, +0.06 with 705 animals in average. Genomic selection did not improve accuracy of genetic breeding values, even with all animals in the reference population. But some improvements in accuracy were noticed for males with fathers genotyped.

To applied genomic selection in dairy goats some improvement had to be done like adding females with higher genetic potential as dams of bucks or considering another methods than GBLUP to estimate marker effects. Nevertheless this study allowed suspecting quantitative trait locus (QTL) on SCS. It is an interesting result needing using methods of QTL detection to specify the localization of it.
Abbreviations glossary
AI: Artificial Insemination
BLUP: Best Linear Unbiased Predictor
CD: Coefficient of Determination (EBV accuracy)
EDC: Effective Daughter Contribution
DGV: Direct Genomic Value
DRP: DeRegressed Proofs
DYD: Daughter Yield Deviation
EBV: Estimated Breeding Value
FP: Fat Percent in milk
FU: Fore Udder
FY: Fat Yield in milk
GBLUP: Genomic Best Linear Unbiased Predictor
GEBV: Genomic Estimated Breeding Values
LD: Linkage Disequilibrium
PA: Parent Average
popQTL: population of females and males genotyped
popIA: population of AI bucks genotyped
PTA: Predicted Transmitting Ability
PP: Protein Percent in milk
PY: Protein Yield in milk
QTL: Quantitative Trait Locus
RUA: Rear Udder Attachment
SCS: Somatic Cells Score
SNP: Single Nucleotide Polymorphism
TO: Teat Orientation
UFP: Udder Floor Position
UP: Udder Profile
YD: Yield Deviation
## Table of contents

Introduction ................................................................................................................................. 1

I. Background .................................................................................................................................. 2

   I.1. Structure of the reference population ..................................................................................... 2
       I.1.A. Extent of LD ..................................................................................................................... 2
       I.1.B. Levels of relationship within reference population ......................................................... 2

   I.2. Multi breeds or multi populations reference population .......................................................... 3
       I.2.A. Using several populations .................................................................................................. 3
       I.2.B. Using several breeds .......................................................................................................... 4
              I.2.B.a. In dairy cattle .............................................................................................................. 4
              I.2.B.b. In beef cattle .............................................................................................................. 4
       I.2.C. Several models to use multi-breed reference population ................................................... 5

   I.1. Use of female genotypes ......................................................................................................... 5

   I.2. Phenotypes used ....................................................................................................................... 6

II. Material and methods .................................................................................................................. 7

   II.1. Reference population in French Alpine and Saanen goats ......................................................... 7

       II.1.A Description of popQTL and popIA population structures .................................................. 7
       II.1.B Phenotypes available .......................................................................................................... 9
              II.1.B.a. YD and DYD ........................................................................................................... 9
              II.1.B.b. Weight of phenotypes .............................................................................................. 11
       II.1.C Genotypes of the reference population ................................................................................. 12
              II.1.C.a. Genotypes available .................................................................................................. 12
              II.1.C.a. Calculation of LD extent and persistence of LD between breeds ............................. 12
              II.1.C.b. Simulation of bucks genotyped ................................................................................... 12

   II.2. Prediction of GEBV ................................................................................................................ 13

       II.2.A Prediction of GEBV accuracy ............................................................................................ 13
       II.2.B Estimation of GEBV ............................................................................................................ 14
              II.2.B.a. GBLUP ..................................................................................................................... 14
              II.2.B.a. Calculation of accuracy ............................................................................................. 14
              II.2.B.b. Cases studied ............................................................................................................. 14

III. Results and discussion ............................................................................................................... 15

   III.1. Structures of the two populations genotyped ......................................................................... 15

       III.1.A. PopQTL .......................................................................................................................... 15
result could indicate that these males are homozygote for a QTL concerning SCS. But further analyses are needed.
Figure 1. Scheme of cross-validation method to estimate accuracy of genomic breeding values
Introduction

Arrival of genomic techniques in dairy cattle selection is a revolution. Indeed the breeding value of an animal could be predicted at birth just using genotype which is a genome scan using dense markers. Genomic selection leads to decrease generation interval in male pathway and increase accuracy of genomic breeding values (GEBV) predictions. Besides, costs of selection are lower than the ones necessary for classical selection because bulls are used earlier and for a shorter time than in progeny testing.

The principle of genomic evaluation is to estimate the breeding value of an animal summing the estimated effects of each marker (SNP: single nucleotide polymorphism or haplotypes, sequence of marker alleles on a chromosome). Marker effects are generally estimated with genomic best linear unbiased prediction model (GBLUP), estimating simultaneously marker effects and random polygenic ones.

A reference population is needed to predict SNP effects and to estimate genomic breeding values (GEBV) of animals without genotypes. The reference population is split in two parts: a training set consisting of old animals with phenotypes and genotypes and a validation set of the most juvenile animals in which only genotypes are considered. The genomic prediction equation is defined from a training set. Then the GEBVs of the validation set are predicted from these equations. Accuracy of genomic predictions is estimated using cross-validation method and is computed from the correlation between GEBVs and the observed phenotypes (Figure 1). The quality of predictions depends on number phenotypes, heritability, number of genetic markers and reference population size.

In France, genomic selection is already done in dairy cattle since 2010 and is under development in dairy sheep. In goats, arrival of 50k chip and recent genotyping allow studying the feasibility of genomic selection in this species. Gains in generation interval with applying genomic to goats and sheep are expected to be lower than in bovine. Increase in accuracy of GEBVs depends on reference population (size and structure) and on relationships between reference population and candidate ones. Therefore aim of this study is to evaluate changes in accuracy using genomic selection in goats (Saanen and Alpine). First, the way to constitute this reference population will be studied. Examining several parameters like structure and linkage disequilibrium (LD) extent in the reference population chosen will lead to predict GEBV accuracy. It will be compared to ones calculated using GBLUP model. This will allow characterizing feasibility and defining the limits of a genomic selection in goats.
I. Background

I.1. Structure of the reference population

When animals are highly related within a population, level of linkage disequilibrium (LD) between markers and quantitative trait loci (QTL) is high (Robert-Granié et al., 2011). GEBVs accuracies are dependent on LD extent and therefore on population structure.

I.1.A. Extent of LD

Linkage disequilibrium between marker and QTL refers to non-random association between alleles of marker and QTL. In genomic selection, LD between marker and QTL has to be persistent in all the reference population across generations (Hayes, 2011).

The deviation from a random association between markers is a measure of LD. If we consider two markers A with alleles A1, A2 and B with alleles B1, B2, D can be calculated as:

\[ D = p(A_1B_1)p(A_2B_2) - p(A_1B_2)p(A_2B_1) \]

where \( p(A_1) \) is the frequency of \( A_1B_1 \) haplotype in the population. The \( r^2 \) (Hill and Robertson, 1968) measure of DL less dependent on allele frequencies is used to compare LD extent between population:

\[ r^2 = \frac{D^2}{p(A_1)p(A_2)p(B_1)p(B_2)} \]

where \( p(A_i) \) is the frequency of allele \( A_i \) in the population. Values of this measure range from 0 for no LD to 1 for complete LD between two markers. If we consider a marker and a QTL, \( r^2 \) is the measure of variation caused by alleles at QTL which is explained by marker (Hayes, 2011). The \( r^2 \) coefficient between two markers at the same distance in two breeds can have the same value even if phases are reversed.

Correlation between two \( r \) coefficients (\( \sqrt{r^2} \)) in two different populations is preferred as a measure for LD phases persistence between populations. The value of \( r \) depends on marker distance and breed considered. Indeed for high marker distance (superior to 100 kb), LD level is the same whatever breeds studied. The largest difference of LD pattern between two bovine breeds (New Zealand Holstein Australian Angus) is obtained for low marker distance (70-75 kb) (De Roos et al., 2008). LD extent is higher in multi-breed reference population than in purebred one. Moreover, persistence of LD phase was less consistent if the training and validation population are very different (Toosi et al., 2010).

I.1.B. Levels of relationship within reference population

Increasing level of relationship between training set and validation set improve GEBV accuracies. Including in training set father, half sibs and full sibs of animals of validation set increase accuracies of 50% for milk yield compare to including non relatives (Habier et al., 2010). And including sires of cows used in validation set in the training set increases female
GEBV accuracy of 47% for milk yield. Using all sires of predicted bulls in the training set increases GEBV accuracies of 26% in fertility and 12% in protein yield compared to training set excluding sire. In fact, adding sires in the training data leads to estimation of GEBV not only from the LD but also from sire genetic information (Lund et al, 2009).

Regarding lower levels of relationship between training and validation set, including half sibs compared to first cousin improves GEBV accuracy of 112% (from 0.27 to 0.57) in Merinos beef sheep (Clark et al, 2012). Even adding maternal and paternal grandsires in the training test increases accuracies of 9.8% in protein yield and of 16.1% in somatic cell score compared to training set without relatives of sires (Pérez-Cabal et al, 2010).

Reliabilities of direct genomic values (DGV), sum of all marker effects excluding polygenic effects and information from non-genotyped relatives, are 88% higher with an increase of 0.55% in average squared relationship between training and candidate sets. These differences observed in reliabilities are proportional of heritability of the trait. But, DGV reliabilities decrease with the level of relationship between animals within reference population. When cows within training set are highly related (descendant of only 5 sires), reliabilities of DGV are 18.3% under DGV reliabilities of low related population (40 sires) for a simulated trait (h²=0.3) (Pszczola et al, 2012).

Strong LD between candidates and reference population caused by high level of relationship between the two could be damaged by recombination at next generations. It seems to be better to genotype and phenotype relatives of candidates like their dam and sire or their sire and maternal grandsire (Robert-Granié et al, 2011). So it is important to pay attention to the structure of the reference population to make genomic selection.

I.2. Multi breeds or multi populations reference population

Multi-breed or several populations in a same reference population increase the size of the reference population and GEBV accuracies but could not be a good choice in terms of LD extent for very divergent breeds or populations.

I.2.A. Using several populations

Using the “European across countries population” as training set for Holstein cows and only one country population as validation, reliabilities of DGV are higher than using only one country in training set. DGV Reliabilities are improved of 21, 11%, 8% and 9% respectively for German, Nordic, Dutch/Flemish and French population compared to national reference population (Lund et al, 2010; Lund et al, 2011). A training set consisted of one population (Danish Red) to predict other ones (Swedish and Finnish Red) does not estimate
well DGV reliabilities (-14% and -17% compare to national evaluation) in dairy traits. Nevertheless the use of combined reference population improves DGV reliabilities (3% for Danish and 7% for Finnish and Swedish) compare to using national reference population. However adding a third population to a two-population reference one does not really increase reliabilities even if increasing is higher for genetically close populations like Swedish and Finnish Red cows (Brøndum et al, 2011).

Adding 1 000 individuals of a simulated population B to a training set made up with 1 000 animals from simulated population A induces an increase on DGV reliabilities of population A of 0.12 (maximum) depending on LD extent and time of divergence simulated. Using two genetically divergent populations by increasing number of divergent generations of 270 between A and B decreases reliabilities of 44.6% (De Roos et al, 2009). Small increase in DGV and GEBV reliabilities with combined reference population is conditioned by relationships between populations, heritability of the trait and marker density. Moreover adding a breed specific effect in the model which estimates GEBV from multi-breed reference population does not led to any improvement (Makgahlela et al, 2011).

I.2.B. Using several breeds

I.2.B.a. In dairy cattle

Using a multi-breed reference population has a small effect on GEBV accuracy predictions whatever the dairy trait considered (Hayes et al, 2009a; Pryce et al, 2011). In Jersey cows, accuracy for fat percent is 1% higher using a multi-breed than purebred reference population and 13% higher compare to non genomic accuracy (Hayes et al, 2009a). A three-breeds reference population induces a small increase in GEBV accuracy compare to using two breeds. Accuracy obtained in milk yield with the Holstein-Jersey-Fleckvieh reference population is 5% higher than the one calculated with Holstein-Fleckvieh reference population. The accuracy gain obtained for small breeds is not really noticed in large breed. In order to predict genomic gain in Holstein breed using only Holstein or multi-breed (Jersey, Fleckvieh, Holstein) reference populations gives the same accuracies of 0.41 in protein and fat yields (Pryce et al, 2011).

I.2.B.b. In beef cattle

Purebred training population (Angus) used with multi-breed validation population (Carcass Merit Project including Angus) gives good predictions for all breeds considered. Correlations between predicted and true genotypic values are 0.946, 0.969, and 0.972 for Hereford Brahman and Angus respectively. Multi-breed reference population increase
correlations of 1% for simulated phenotypes compare to Angus purebred reference population. Nevertheless correlation between predicted genetic merit and phenotype is better using multi-breed reference population (Kizilkaya et al, 2009).

Training in purebred population to predict genetic merit of multi-breed animals is possible but induce small increase in GEBV accuracies especially for large breeds. It is useful in livestock because of crossbreed or multi-breed lines but not in dairy cattle because accuracies are improved only in small breeds. GEBV are more accurate with a multi-breed than with purebred reference population in case of population without individuals genotyped and phenotyped (Pryce et al, 2011). Results using several breeds in a unique reference population were summed up in Annexe A.

I.2.C. Several models to use multi-breed reference population

Several statistical models tend to predict GEBV of animals like Bayesian models (Bayes A, B, etc) or GBLUP. At this time no any model is assume to be better than others. In genomic evaluation with multi-breed (Holstein-Jersey-Fleckvieh) reference population, GEBV accuracies in dairy traits calculated with GBLUP model (0.4) are really close to the ones calculated with BayesA (0.37) (Pryce et al, 2011). Some studies have shown the superiority of Bayesian model in case of multi-breeds reference population. Bayes A and Bayes_SSVS are more accurate than GBLUP for GEBV of Jersey animals using Holstein-Jersey reference population (Pryce et al, 2011). It seems to be better (16% higher than BLUP on weight at ultrasound scanning) to predict the breed mostly present (Merinos sheep) in the reference population (Daetwyler et al, 2010). On cows fat percent GEBV are 12% higher with Bayesian (0.82) models than with GBLUP (0.73). But BLUP is better for large breed like Holstein cows (Hayes et al, 2009a). GBLUP model with a specific breed effect leads to higher DGV reliabilities than Bayesian models (Makgahlela et al, 2010). A model with SNP effects different but correlated between breeds is 4.1% more accurate than model with same SNP effects for genetically distant breeds (Varona et al, 2010).

A reference population combining populations could improve the reliability of predictions except if marker effects are very different between populations. Markers need to be very close to QTL, the only ones in persistent LD with QTL across breeds or populations. So many projects focus on high density chip like the 800 k chip in ANR GEMBAL project for French beef cattle.

I.1. Use of female genotypes

Since 2011, decrease of genotyping cost allows examining the interest of female genotypes like ANR AMASGEN project (with a low density chip) which is expected to increase genetic
Figure 2. Diagram of breeding scheme simulated in Mc Hugh et al (2011) study

Figure 3. Annual genetic gain for several numbers of genotyping per year (0, 250, 1 000, 2 000) and several proportions of males and females in the reference population (Sørensen and Sørensen, 2010)
gain and decrease inbreeding rates (Moser et al., 2010). In average combined male and female genotypes increases GEBV accuracies of 25% compare to progeny testing (Schaeffer, 2006). DGV reliabilities of cows using only female genotypes are 39% higher than EBV on energy balance (Verbyla et al., 2010) whereas genetic gain is more than twice higher using female genotypes in genomic selection than in progeny testing (McHugh et al., 2011). In a simulated dairy cattle breeding scheme (Figure 2), using female genotypes increases the annual genetic gain compare to genomic selection with only male genotypes of 44% (McHugh et al., 2011) or 21% (Buch et al., 2010) depending on simulations. Using female genotypes to preselect them at birth leads to genetic progress 2.4% low than genomic selection which preselect males at birth (Schrooten et al., 2005).

Moreover increase number of cow genotypes used in reference population (from 1 000 to 2 000) improves DGV reliabilities of 40% (0.3 compare to 0.18) on a simulated trait (h²=0.3) (Calus et al., 2011). Indeed use 10% of female genotypes compare to 2% leads to an increase in accuracy of 4% for a simulated trait (h²=0.25). Furthermore annual genetic gain is maximum for 25% of male genotypes in reference population (Figure 3). In this case, increasing proportion of female genotypes in reference population (from 75% to 100%) does not improve the annual genetic gain (Sørensen and Sørensen, 2010). Using top ranking female genotypes does not improve GEBV accuracies compare to female genotypes chosen at random (Jiménez-Montero et al., 2010). Using female genotypes in the reference population is a good way to accelerate genetic progress and to improve accuracy of GEBV of both sexes. It is interesting because genotyping females would lead to better mating and high merit calves that would compete with current sires (Sørensen and Sørensen, 2010). Nevertheless even if female genotypes are not used, applying selection intensity (multiple ovulation and embryo transfer) in dams of cows leads to an increase in genetic gain (Pryce et al., 2010). Results are summed up in Annexe B.

I.2. Phenotypes used
Phenotypes mostly used for bulls in genomic selection are daughter yield deviations (DYD). DYD are average performance of bull daughters corrected for environmental effects and for their dam genetic merit. For females, yield deviations are used as phenotypes, and defined as $YD = y - (\beta + p)$ where $y$ is yield of cow, $\beta$ is fixed effects and $p$ is permanent environment effect (Szyda et al., 2008). If DYD are not available like for stranger bulls, deregressed proofs (DRP) are considered as phenotypes (Robert-Granié et al., 2011). The DRP are average performance of daughters calculated from genetic merit of bulls using the inverse of a genetic evaluation method. In Nordic dairy cattle using DRP with single step genomic
evaluation improve DGV reliabilities of 6.7%, 12.9% and 10% for milk, protein and fat yields respectively compare to using DYD (Mäntysaari et al, 2011).

Estimated breeding values (EBV) which already contain genetic information can be used as phenotypes. With GBLUP, GEBV reliabilities are slightly higher (+0.7%) using EBV than using DYD. However this difference depends on heritability (Guo et al, 2010). EBV easier to use because directly obtained from genetic evaluation contains smaller random error than DYD. Nevertheless EBV includes information of all relatives leading to double counting because predictions of GEBV also include information from relatives. DRP could be a good alternative because it does not take into account relatives and are easier to obtained compare to DYD (Guo et al, 2010).

A “single step approach” is for the moment the only method which analyses directly performance records for genomic predictions (Legarra et al, 2011). Indeed this approach combines phenotypic, genomic and pedigree information into a single simultaneous analysis. Using a one step approach leads to similar accuracy and smaller values (-73%) for prediction error variance and mean square error than using DYD in classical genomic evaluation method (Vitezica et al, 2010). Nevertheless based on coefficients of determination, one-step approach is 2.5% less accurate than multiple-step one using DYD (Aguilar et al, 2010).

II. Material and methods

The aim of this study is to characterize the feasibility of genomic selection in the French dairy goat population. The interest of genomic selection consists in the decrease of generation interval in a higher accuracy of breeding values. For this study the feasibility of genomic selection was estimated by comparing GEBV’s accuracies to EBV ones. Two populations of genotyped animals (popIA and popQTL) were considered as the two sets of a reference population.

II.1. Reference population in French Alpine and Saanen goats

As described above, population structure has an important effect on accuracy of GEBV. Therefore the two populations of genotyped animals were first analyzed in terms of pedigree and genetic structure.

II.1.A Description of popQTL and popIA population structures

PopQTL and popIA were the two populations genotyped with the Illumina 50K goat BeadChip. The popQTL was a population initially dedicated to QTL detection. It consisted in 20 AI bucks (9 Saanen and 11 Alpine) born between 1999 and 2004 and their 2,246 daughters with an average of 112 daughters genotyped per male (from 74 to 127). Bucks of popQTL
population had high total merit index (ICC, including type and production traits) from 0.97 to 5.76. The popIA population consisted in 1 012 AI bucks (429 Saanen and 582 Alpine) including bucks from popQTL, born between 1985 and 2011 with an ICC merit from - 6.48 to 7.78. Among these 1 012 bucks, only the 850 youngest bucks born after 1998 will be genotyped. Negatives ICC values were obtained for former bucks because indexes were expressed as deviation from a rolling base updated each year with current bucks (with a better genetic level). At this time only 85 genotypes were available including 39 Saanen bucks and 46 Alpine ones.

The description of the population consisted in analysis of pedigree and genetic structure. The pedigree structure of the two populations was designed with Pedigree viewer software (Kinghorn and Kinghorn, 2007) considering genotyped animals and 4 anterior generations (all parents known). Analysis of genetic structure was done with Pedig software (Boichard, 2006) considering 26 generations. The quality of pedigree information, given by the number of equivalent generation known, was evaluated using option “Ngen” of Pedig software:

\[
N_{\text{gen}} = n_p + n_{gp} \left( \frac{1}{2} \right) + n_{ggp} \left( \frac{1}{2} \right)^2 + n_{ggp} \left( \frac{1}{2} \right)^3 + \cdots + n_a \left( \frac{1}{2} \right)^n
\]

where \( n_p \) is the number of parent known, \( n_{gp} \) the number of grandparents known, \( n_{ggp} \) the number of great grandparents known and \( n_a \) the number of ancestor known at generation \( n \) (Danchin-Burge, 2011). The average number of ancestors by individual was also computed. Average inbreeding and parental coefficients were calculated independently for all animals: 1) from popQTL AI bucks pedigree, 2) from their daughter pedigree and 3) from popIA bucks pedigree considering the whole pedigree including 26 generations at maximum. The popQTL pedigree consisted in 34 771 animals, 18% had at list one parent unknown and 14% with both parents unknown. 17 912 animals made up the popIA pedigree, 21% had at least one parent unknown and 18% with both parents unknown. Inbreeding coefficient, calculated with “meuw” option of Pedig software, is the probability at a given locus that an individual received similar alleles from his both parents. It could be expressed as

\[
F_i = \left( \frac{1}{2} \right)^n \times (1 + F_A)
\]

where \( n \) is the number of individuals on parental line and \( F_A \) is inbreeding coefficient of common ancestor (Verrier et al, 2001). Average parental coefficients had been calculated with “par” option of Pedig software taking 50 samples of 30 individuals at random. Parental coefficient between I and J is the probability that two alleles of a gene are similar in I and J genome (Verrier et al, 2001). Effective number of founders was calculated for AI bucks of popQTL and popIA with “probability of gene
Figure 4. Scheme of the two cases for achievement of phenotypes: one with all females’ information and another only with popQTL females’ information to analyze interest of female genotypes

- Accurate YD of females from popQTL
- Accurate DYD_{allfem} of the 685 bucks chosen

Adding female genotypes is not expected to actually improve GEBV accuracy of males

- Less accurate YD of females from popQTL
- Less accurate DYD_{femg} of the 20 sires of females from popQTL

Adding female genotypes to genotypes of their 20 sires is expected to highly increase level of bucks GEBV accuracy

Figure 5. The five morphology position used for this experiment as morphology phenotypes
origin” option of Pedig software. This effective number of founders is the inverse of the probability that two genes in the population come from the same ancestor: this is the number of founders that given the same genetic variability if they all contribute at the same level (De Rochambeau et al., 2003).

II.1.B Phenotypes available
Phenotypes mostly used in genomic evaluation are DYD for males and YD for females. They are derived from classical genetic evaluation. Figure 4 is a scheme of methods used to obtain DYD and YD in this study. Genetic evaluations did not include only animals of popIA and popQTL populations. DYD and YD were more accurate than expected because animals have already progeny and relatives in the all population. Therefore two classical genetic evaluations were made to obtain two DYD or YD per animal.

II.1.B.a. YD and DYD
The two evaluations were: (i) an evaluation with all official females (2,885,244 individuals), DYD_{allfem} were well known because all bucks daughters were included and, (ii) an evaluation on the 2,246 females from popQTL only, DYD_{femg} were not as accurate as in the first case. This second case was similar to a situation where only data from females were available to predict young males (not progeny tested) breeding values. In this case including females genotyped was expected to improve GEBV accuracies. For this case genetic evaluation model was simplified without heterogeneity of variance and permanent environment because of few data. Level of fixed effects like year, region and parity were pooling to allow estimation of effects (Annexe C). To underline the interest of female genotypes, a genomic evaluation using phenotypes obtained in the second case (DYD_{femg} and YD) and male and female genotypes was compare to another one using only male phenotypes (DYD_{femg}) and genotypes. YD of genotyped females and their deviations were calculated with preadj option of GENEKIT software (Ducrocq, 1998). DYD were derived as
\[ DYD = \frac{\sum_{i=1}^{n} d_{wi}}{\sum_{i=1}^{n} w_{i}} \] where i is the number of observations (lactations of all daughters) per male, devi deviation of the observation i and wi is weight of observation i. Deviation of observation was performance corrected for non genetic effect and for half dam genetic merit. In this study traits considered were milk yield, fat yield (FY), protein yield (PY), fat percent (FP), protein percent (PP), somatic cell score (SCS) and the five type traits used in combined morphologic index (ICM): fore udder (FU), udder profile (UP), udder floor position (UFP), teat orientation (TO), and rear udder attachment (RUA) (Figure 5).
Table 1. Phenotypes obtained with genetic evaluation on all official females (2,885,244 goats) YD for females (n=1,946) and DYD for males (n=705).

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Mean</th>
<th>Standard error</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YD</td>
<td>DYD</td>
<td>YD</td>
<td>DYD</td>
</tr>
<tr>
<td>Milk</td>
<td>197</td>
<td>76.7</td>
<td>128</td>
<td>40.6</td>
</tr>
<tr>
<td>FY</td>
<td>7.18</td>
<td>2.99</td>
<td>4.45</td>
<td>1.51</td>
</tr>
<tr>
<td>PY</td>
<td>6.93</td>
<td>2.83</td>
<td>3.56</td>
<td>1.12</td>
</tr>
<tr>
<td>FP</td>
<td>0.74</td>
<td>0.52</td>
<td>2.87</td>
<td>1.37</td>
</tr>
<tr>
<td>PP</td>
<td>1.10</td>
<td>0.59</td>
<td>1.53</td>
<td>0.73</td>
</tr>
</tbody>
</table>

FY: Fat Yield, PY: Protein Yield, FP: Fat percent, PP: Protein percent

Table 2. Phenotypes obtained with genetic evaluation only on the first lactation of genotyped females (2,246 goats) YD for females and DYD for males

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Mean</th>
<th>Standard error</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Correlation with official evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YD</td>
<td>DYD</td>
<td>YD</td>
<td>DYD</td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>-2.89</td>
<td>1.39</td>
<td>-807</td>
<td>619</td>
<td>0.92</td>
</tr>
<tr>
<td>FY</td>
<td>-4.32</td>
<td>2.36</td>
<td>-36.7</td>
<td>19.8</td>
<td>0.92</td>
</tr>
<tr>
<td>PY</td>
<td>-1.24</td>
<td>4.14</td>
<td>-27.4</td>
<td>16.9</td>
<td>0.92</td>
</tr>
<tr>
<td>FP</td>
<td>-3.04</td>
<td>2.94</td>
<td>-14</td>
<td>15.9</td>
<td>0.94</td>
</tr>
<tr>
<td>PP</td>
<td>-0.61</td>
<td>1.64</td>
<td>-5.68</td>
<td>6.48</td>
<td>0.94</td>
</tr>
</tbody>
</table>

SCS: Somatic Cell Score, FU: Fore Udder, UP: Udder Profile, UFP: Udder Floor Position, TO: Teat Orientation, RUA: Rear Udder Attachment

Table 3. Genetic parameters used to compute EBV in official evaluations in Alpine and Saanen goats

<table>
<thead>
<tr>
<th>Trait</th>
<th>Genetic variance</th>
<th>Residual error variance</th>
<th>Permanent environment variance</th>
<th>Heritability</th>
<th>Genetic variance</th>
<th>Residual error variance</th>
<th>Permanent environment variance</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>11666</td>
<td>19443</td>
<td>7777</td>
<td>0.3</td>
<td>The same as in Alpine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FY</td>
<td>14.26</td>
<td>23.76</td>
<td>9.50</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PY</td>
<td>9.04</td>
<td>15.06</td>
<td>6.03</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>8.73</td>
<td>5.24</td>
<td>3.49</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td>2.32</td>
<td>1.39</td>
<td>0.93</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCS</td>
<td>0.30</td>
<td>0.85</td>
<td>0.39</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FU</td>
<td>0.24</td>
<td>0.71</td>
<td>None</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UP</td>
<td>0.72</td>
<td>1.29</td>
<td>(because only one observation by female)</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UFP</td>
<td>0.27</td>
<td>0.51</td>
<td>0.34</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TO</td>
<td>0.23</td>
<td>0.47</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUA</td>
<td>0.37</td>
<td>1.05</td>
<td>0.26</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- Table 1 shows statistical results on YD and DYD given for animals genotyped and calculated with evaluation on all females. In this case YD of females genotyped were on average of 197.2 ± 256 for milk yield and 0.74 ± 5.8 for fat percent (FP). Average DYD was 76.7 ± 80 for milk yield and 0.52 ± 2.8 for FP. DYD of all males from popIA and popQTL were given for the 860 bucks with daughters in production in 2012. The average number of observations (lactations of daughters) for DYD was 801 ± 2 388 (from 1 to 10 120) per buck.

- Table 2 presents statistical results on YD and DYD obtained with evaluation on the 2 246 females genotyped and given for the 2 266 animals from popQTL. Average YD was - 2.89 ± 362 on milk yield and - 3.04 ± 5.8 on FP. DYD was on average - 0.56 ± 102 on milk yield and 0.15 ± 2.5 on FP. Table 2 also shows correlations between DYD (and YD) obtained with evaluation on all females and those obtained with evaluation on females genotyped. These correlations ranged from 0.75 to 0.94, because model of non official genetic evaluations were different and variance heterogeneity was not taking into account. YD and DYD values in official method were not comparable with the one obtained for evaluation on females genotyped. Nevertheless minimum and maximum of these YD and DYD ranged in the same order than those obtained on all official females.

EBVs for Alpine and Saanen females were calculated together for dairy traits. It was known that there were slight differences between their performances (Bélichon et al, 1999). In this study differences between Saanen YD and Alpine one in the whole population (2.8 millions) were noticed. YD for milk yield were in average of - 32.39 ± 372 in Alpine and 30.35 ± 380 in Saanen (results not shown). Phenotypes of Saanen goats were in average superior than the ones in Alpine for milk protein and fat yield but it was the contrary for fat and protein percent. These differences induced taking into account a fixed breed effect in genomic evaluation model. Standard errors of YD in Alpine and Saanen were similar. So heterogeneity of variance had not to be implemented in the genomic evaluation model. SCS: Somatic Cell Score, FU: Fore Udder, UP: Udder Profile, UFP: Udder Floor Position, TO: Teat Orientation, RUA: Rear Udder Attachment

Table 3- Table 3 presents genetic parameters for several traits, only type traits and SCS are different between the two breeds. Higher heritability is 0.4 for FP and protein percent (Spangler et al, 2008), 0.3 for milk yield, FY and PY, 0.19 for SCS in Alpine and 0.21 in Saanen and from 0.26 to 0.36 for type traits. YD of Saanen and Alpine for those functional traits were calculated separately.
Table 4. YD (n=1,946) and DYD (n=705) calculated with all females (2.8 millions) for Somatic cell Score and type traits considering Alpine and Saanen together.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean YD</th>
<th>Standard error YD</th>
<th>Minimum YD</th>
<th>Maximum YD</th>
<th>Mean DYD</th>
<th>Standard error DYD</th>
<th>Minimum DYD</th>
<th>Maximum DYD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCS</td>
<td>0.117</td>
<td>0.026</td>
<td>0.99</td>
<td>0.26</td>
<td>-2.75</td>
<td>-0.82</td>
<td>3.74</td>
<td>0.81</td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FU</td>
<td>0.008</td>
<td>-0.007</td>
<td>0.93</td>
<td>0.30</td>
<td>-2.59</td>
<td>-1.22</td>
<td>3.83</td>
<td>2.25</td>
</tr>
<tr>
<td>UP</td>
<td>-0.127</td>
<td>0.007</td>
<td>0.87</td>
<td>0.29</td>
<td>-3.70</td>
<td>-1.89</td>
<td>2.59</td>
<td>1.27</td>
</tr>
<tr>
<td>UFP</td>
<td>-0.093</td>
<td>0.013</td>
<td>1.32</td>
<td>0.39</td>
<td>-5.09</td>
<td>-2.67</td>
<td>3.69</td>
<td>1.99</td>
</tr>
<tr>
<td>TO</td>
<td>-0.094</td>
<td>0.004</td>
<td>0.88</td>
<td>0.26</td>
<td>-3.09</td>
<td>-1.20</td>
<td>5.03</td>
<td>1.15</td>
</tr>
<tr>
<td>RUA</td>
<td>-0.123</td>
<td>0.027</td>
<td>1.37</td>
<td>0.44</td>
<td>-4.14</td>
<td>-2.86</td>
<td>5.54</td>
<td>2.69</td>
</tr>
</tbody>
</table>

Figure 6. Distribution of EDC values on SCS for the 685 males in the reference population.

Figure 7. Graph of EDC values on milk yield function of the number of daughters in official genetic evaluation for the 685 males in the reference population.

Table 5. Correlations between EDC values and number of daughters in official genetic evaluation.

<table>
<thead>
<tr>
<th>Traits</th>
<th>Milk</th>
<th>FY</th>
<th>PY</th>
<th>FP</th>
<th>PP</th>
<th>SCS</th>
<th>Type traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ</td>
<td>0.990</td>
<td>0.989</td>
<td>0.990</td>
<td>0.984</td>
<td>0.986</td>
<td>0.996</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Table 6. Average weights of male phenotypes (EDC) for evaluation with all females (official) and evaluation with only females genotyped (non official).

<table>
<thead>
<tr>
<th>Traits</th>
<th>Milk</th>
<th>FY</th>
<th>PY</th>
<th>FP</th>
<th>PP</th>
<th>SCS</th>
<th>AVP</th>
<th>PRM</th>
<th>PLA</th>
<th>ORT</th>
<th>AAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>All females</td>
<td>413.5</td>
<td>413.2</td>
<td>413.5</td>
<td>389.2</td>
<td>389.3</td>
<td>159.8</td>
<td>117.9</td>
<td>118.3</td>
<td>117.4</td>
<td>117.9</td>
<td>117.5</td>
</tr>
<tr>
<td>Females genotyped</td>
<td>88.8</td>
<td>88.7</td>
<td>88.8</td>
<td>94.8</td>
<td>95.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4 shows statistical results on YD and DYD given for animals genotyped for SCS and type traits obtained with evaluation on all females. On all females genotyped only 2 233 had been registered for morphology. YD obtained for type traits were lower than those obtained for milk traits. In average all YD on type traits were around - 0.1 except for fore udder where average YD is close to 0.01. DYD were in average from - 0.007 to 0.027, the highest DYD was obtained for rear udder attachment and the smallest one for fore udder. For somatic cell score only 2 229 females had been recorded and 789 males had daughter with SCS records. Average YD in SCS were 0.117 ± 2 and average DYD were 0.03 ± 0.52 which was in the same order of magnitude than the ones for type traits.

II.1.B.b. Weight of phenotypes

Weight of female phenotypes taken here was weight of lactation in the official evaluation, 1 for fist lactation and 0.8 for second and third lactation in goats. As in dairy cattle weight of male performances considered here was the effective daughter contribution (EDC). It took into account almost all the above-mentioned factors and was calculated as:

$$EDC_i = \sum_k \frac{\lambda R_k}{4 - R_k(1 + R_{\text{dam}})} \quad \text{with} \quad R_k = \frac{n_k h^2}{1 + (n_k - 1)r}$$

where $R_k$ is the reliability of daughter k’s performance and $R_{\text{dam}}$ reliability of her dam performance, $n_k = \sum_l 1 - \frac{1}{n_{jkl}}$ where $n_k$ is number of lactations of daughter k corrected for contemporary group and $n_{jkl}$ is contemporary group size in which daughter k made its $l^{th}$ lactation, and $\lambda = \frac{4 - h^2}{h^2}$ where $h^2$ is the heritability of the trait and r the repeatability of a record (Fikse and Banos, 2001). In this study EDC were calculated using crEDC software (Sullivan, 2010) and separately for the two breeds for SCS and type traits (Table 3). Figure 6 presents the distribution of EDC values for SCS identical to EDC distribution for other traits, almost following Gaussian distribution. Figure 7 is a representation of EDC values linearly proportional to number of daughters in official genetic evaluation. The bucks with less than 100 daughters were the young ones born after 2007. Table 5 shows high correlations (from 0.99 to 0.98) between EDC and number of daughters in official evaluation.

Table 6 presents average EDC calculated with evaluation on all females and on only females genotyped. EDC in official case (413 on FP) were larger than the one obtained only with female genotypes (89 on FP) and were ten times weight of female phenotypes.
Figure 8. Scheme of calculation method of $r^2$ measure of LD within a population
II.1.C  Genotypes of the reference population

II.1.C.a. Genotypes available

In this study 49,674 SNP out of 53,347 synthetized on the Illumina 50K goat BeadChip were validated after quality control. Control consisted in a call freq (percent of animals with assigned genotype per SNP) superior to 0.99, a call rate (percent of SNP available per animal) superior to 0.98, a minor allele frequency of 1%, Hardy Weinberg equilibrium control and filiation. At the end, 2,039 genotypes were available on the 2,351 animals genotyped with 1,946 female and 9 male genotypes (on 2,255) from popQTL and all (84/85) from popIA were available including the 11 bucks in common to popQTL and popIA.

This chip seemed to be sufficient to achieve accurate genomic selection according to (Hayes and Goddard, 2010). Indeed, they suggest that 10 NeL (where Ne is the effective population size and L the genome length in Morgans) i.e. 36,000 markers were necessarily to predict accurate GEBVs. At the start of the present study no genetic or physical map was available for the goat SNP markers. SNP were mapped on 2,316 scaffolds with length from 1 base pair (bp) to 9.93 Mega base (Mb) including from 1 to 243 markers for the largest scaffold.

II.1.C.a. Calculation of LD extent and persistence of LD between breeds

Estimation with genomic data of LD persistence between breed and LD extent in the reference population was calculated with a Fortran software developed in SAGA lab and modified to consider largest scaffolds instead of chromosomes. $r^2$ were derived for haplotypes two by two even for non consecutive SNP for each scaffold (Figure 8) knowing distance between the two SNP considered. Average $r^2$ values considering as LD extent were calculated by regular distance interval of 0.02 Mb which allowed printing $r^2$ of LD in function of marker distance.

For the persistence of LD phases between Saanen and Alpine, average correlation of $r$ values between Alpine and Saanen were derived by interval distance of 0.02 Mb.

II.1.C.b. Simulation of bucks genotyped

At the start of this study only 85 buck genotypes were available, but at the end of the year 768 additional bucks will be genotyped. To estimate the interest of genomic selection in our goat population, we had to simulate buck genotypes to be as close as possible to a real genomic evaluation. Ideally buck genotypes would be simulated considering phenotypes, pedigree and LD information. But such a simulation is highly time consuming. In this study, the 612 buck genotypes (born before 2002) from popIA were simulated using Memestra software (Guillaume, 2010). It allowed simulating genotypes conditionally to pedigree (on
three generations including animals already genotyped), allele frequency in the reference population, and to LD between markers using information from SNP position on genome. Because information of SNP position was only known on scaffolds all simulations were done by scaffold (2,316). Allele frequencies were calculated in the reference population (1,955 animals from popQTL and 84 from popIA).

II.2. Prediction of GEBV

II.2.A. Prediction of GEBV accuracy

To identify if the reference population used in this study will lead to good predictions, several equations predicting accuracy functions of heritability and reference population size were used. We have considered three expressions: (i) \[ r^2 = \sqrt{1 - \frac{\lambda}{2N\sqrt{a}} \times \ln \left( 1 + a + 2\sqrt{a} \right) \left( 1 + a - 2\sqrt{a} \right)} \]
with \[ a = 1 + \frac{\lambda}{N}, \quad \lambda = \frac{2NeLK}{h^2} \] and \[ k = \frac{1}{\log(2Ne)} \] where L is the genome length in Morgan. It assumes that accuracy of estimation of QTL effect is higher for QTL with intermediate allele frequency because they explain more the genetic variance (Goddard, 2009).

(ii) Daetwyler et al (2010) propose a function developing for case-control studies assuming that all estimations of QTL effects have the same accuracy: \[ r^2 = \frac{Nh^2}{Nh^2 + (2NeL)} \]

(iii) Meuwissen (2009) proposes this equation: \[ r^2 = \frac{Nh^2}{Nh^2 + 4NeLv} \] where \( 4NeLv \) is the number of QTL loci on genome. It can be approximate by \[ r^2 = \frac{Nh^2}{Nh^2 + \frac{2NeL}{\ln(4Ne)}} \]

To analyse the \( r^2 \) parameters, we have considered an effective population size of 250 animals (III.1.B), and a length of genome of 30 Morgans. These equations were derived only using male genotypes. As we showed in the background part, including female genotypes had not the same weight than including male ones in terms of GEBV accuracy. Female genotypes were considered as equivalent to 30% (INRA internal communications) of male genotypes, so the 1,946 females in the reference population were assumed to be equivalent to almost 580 male genotypes.

Two reference population sizes were compared. The first one includes only 705 male genotypes (84 from popIA + 9 from popQTL + 612 simulated) and the second one 1,300 (705 + 580) animals. Finally, a trait with \( h^2=0.3 \) (like milk yield) was chosen to evaluate the \( r^2 \) parameters.
II.2.B. Estimation of GEBV

II.2.B.a. GBLUP

To estimate GEBV of both females and males the genomic linear unbiased predictor (GBLUP) using genomic BLUPF90 (Misztal et al., 2011) was used. The BLUPF90 is a software considering a mixed model: \( y = X\beta + Zm + e \) where \( y \) is a vector of phenotypes (YD and DYD), \( X \) is the incidence matrix relating fixed effects (\( \beta \)) to the individuals, \( Z \) is a design matrix allocating observations to breeding values (\( m \)) and \( e \) is a vector of random normal errors. In this model genotypes information is taking into account in genomic relationship matrix \( G \) which is defined as \( G = \frac{ZZ'}{k} \) where \( k \) is a standardization factor which connects marker effects variance to the additive genetic variance. This matrix \( G \) is derived from pedigree and genotypes information. In this study pedigree information was considered as 5% and genotypes information as 95% of the matrix. Concerning phenotypes, male observations was considering to be twice DYD because DYD is derived from \( \frac{1}{2} \) of sire genetic merit, not as animal genetic merit, and had to be multiplied by two to be comparable to female phenotypes.

II.2.B.a. Calculation of accuracy

Estimates of additive genetic values of all animals and standard deviations of this prediction errors were obtained from GBLUP model. An estimation of the accuracy of prediction was \( \sigma^2 = (1-CD) \times \sigma^2 \), where \( \sigma^2 \) is the variance of prediction error, \( \sigma^2 \) the genetic variance of the trait and CD the prediction accuracy. So we could compute the prediction of accuracy by: \( CD = \frac{\sigma^2 - \sigma^2}{\sigma^2} \).

II.2.B.b. Cases studied

The aim of this study was to examine three situations:

i) The first one was to point out the effect of reference population size on accuracy of GEBV. In this study only simulated genotypes were used. Average GEBV accuracies in a reference population with 93 (84 + 9) males (adding or not females) were compared to those in a reference population including 705 males (adding or not females).

ii) Studying interest of female genotypes consisted in comparing GEBV accuracies in two situations: with or without female genotypes. Four cases of reference populations were studied:

- with 20 males phenotyped (with non official evaluation) and genotyped from popQTL (adding or not females),
- with 93 males from popIA and popQTL (with or without females)
- with 93 males from popIA and popQTL with simulated genotypes (with or without females)
**Figure 9.** Scheme of cases studied for analysis of goat genomic selection results where P signify that animals are phenotyped (Poff with official evaluation on 2.8 millions of animals and Pnoff with 2,246 females) and G that animals are genotyped. Animals with only G are young males which are only genotyped with no phenotypes (no daughters).

**Figure 10.** Pedigree of the 20 AI bucks (0) that sired the QTL families representing 5 generations in popQTL (Pedigree viewer©)

**Table 7.** Inbreeding coefficients for pedigree file of IA bucks genotyped animals (males) and of their daughters (females)
- with 705 (93 + 612) males genotypes simulated (with or without females)

iii) Feasibility of genomic selection consisted in comparing GEBV accuracy on these young males to ascendance accuracy \( CD_{asc} = 0.25 \times CD_{dam} + 0.25 \times CD_{ sire} \), it was the expected accuracy of genetic value at birth excluding genomic information.

Several cases studied were summed up to Figure 9. Correlations between genomic solutions and classical BLUP solutions (with all females or with only females genotyped were calculated for each case).

III. Results and discussion

III.1. Structures of the two populations genotyped

III.1.A. PopQTL

Figure 10 shows pedigree scheme for 5 generations of AI bucks sire line. Individuals circled in yellow are the AI bucks; their ancestors are represented in the top of the scheme. As expected, the 20 sires of the QTL design were not directly related. The first degree of relationship between the 20 bucks was first cousin. The number of equivalent generations known, quite high, was 7.89 in this pedigree, around 8 generations known. Average inbreeding coefficient was 2.11% (1.86% for Alpine and 2.40% for Saanen) with a maximum of 25% for AI bucks of popQTL population. A previous study found an average inbreeding coefficient of 2.06 % for Saanen and 1.05 % for Alpine AI bucks born between 1989 and 1998 (Palhière, 2001). The higher coefficients, found in the two breeds, could be explained by a general increase of inbreeding coefficient after 1999 and of relationship between grand-sire of some AI bucks genotyped and groups of half-sibs for females. The average parental coefficient was 2.1% (2.9% for Alpine and 3.6% for Saanen bucks, Table 8), it is close to those obtained between 1989 and 1998 (3.3% for Saanen males and 2.3% for Alpine ones (Palhière, 2001)). Nevertheless inbreeding and parental coefficients were smaller than those obtained for dairy cattle (0.6 for inbreeding coefficient in Holstein (Miglior, 2000)).

Females of popQTL population were all highly related: 35% of females genotyped have at least one full sib in the population. Average inbreeding coefficient was 2.34 % (2.11% for Alpine and 2.63% for Saanen goats) with 51.7% of inbred animals considering the whole pedigree. As for males, females with high inbreeding coefficient are scarce (Figure 11 a b). These result were close to inbreeding coefficients found in Saanen goats (2.1%) and in Alpine goats (2.35%) for all females born between 2006 and 2008 (Danchin-Burge, 2011).
Figure 11. Distribution of inbreeding coefficient for pedigree file of IA bucks genotyped (a) and of their daughters (b)

Table 8. Parental coefficient and effective founders of pedigree file of IA bucks genotyped (males) or of their daughters (females)

<table>
<thead>
<tr>
<th></th>
<th>Average parental coefficient</th>
<th>Standard error of parental coefficient</th>
<th>Number of Effective founders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>0.021</td>
<td>0.042</td>
<td>156</td>
</tr>
<tr>
<td>Females</td>
<td>0.021</td>
<td>0.033</td>
<td>197</td>
</tr>
<tr>
<td>Males</td>
<td>0.018</td>
<td>0.024</td>
<td>271</td>
</tr>
</tbody>
</table>

popQTL | popIA |
Average parental coefficient of genotyped females was equal to 2.1%, similar to genotyped males (1.9% for Alpine and 2.5% for Saanen goats, Table 8). It was higher than the ones found in Saanen (1.4%) and in Alpine (1%) goats born between 1989 and 1999 (Palhière, 2001).

Number of effective founders (Table 8) computed for Alpine and Saanen males was 156: 54 effective founders for Saanen bucks and 117 for Alpine. These results were similar to the ones found for Saanen (50) and for Alpine (119) AI bucks born between 1989 and 1999 (Palhière, 2001).

### III.1.B. PopIA

Pedigree of popIA was better known than pedigree of popQTL. The number of equivalent generation known was 9.3, one more generation known than in popQTL pedigrees. And the average number of ancestors known for an individual (3 836) was more than twice than the one in popQTL. AI bucks of popIA were more related than bucks of popQTL. 132 bucks from popIA had at least one son in the population genotyped with an average of 5 ± 8 sons per buck. PopIA consisted in 885 bucks having a half sib by sire, 110 animals with half sib by dam and 14 bucks with full sib. PopIA and popQTL had a high level of relationship. Indeed popIA included the 20 bucks from popQTL, among which 14 have 79 sons (6 ± 10 per buck in average) in the popIA. Except the 20 bucks including both in popIA and popQTL, their sons and their sires, 2 490 animals (932 males) were common in the two pedigrees: 33% were anterior to the 4th generation of pedigree of popIA. Twelve bucks genotyped (6 Saanen) and born between 1985 and 1999 in popIA were sires of bucks from popQTL. And 7 bucks from popQTL are grand sires of bucks from popIA. The last degree of relationship between the popIA and popQTL was second cousin.

The average inbreeding coefficient of popIA population was 2.17% (1.88% for Alpine and 2.52% for Saanen bucks) (Table 7). The average parental coefficient was equal to 1.8% (1.8% for Alpine and 2.4% for Saanen), lower than the one found in popQTL (2.1%).

The number of effective founders was 271: 214 from Alpine and 87 from Saanen breeds, higher to the one found in males popQTL pedigree because of a new management of inbreeding (selection with minimum relationship method) in goat species since 2002 (Colleau et al, 2007). Selection on AI bucks was managed within family (same maternal grand sire) to maintain genetic progress and decrease bucks inbreeding.
Figure 12. Evolution of effective founders’ number between 1970 and 2012. Effective number of founders is calculated within 4 years of birth.

Figure 13. LD extent in the multi-breed reference population (Alpine and Saanen) function of distance between two markers (Mb)

Figure 14. LD extent in Alpine breed (DLR13) and in Saanen breed (DLR11) function of distance between two markers
The Figure 12 represents the evolution of effective number of founders by year: an increase has been noticed since 2002 with the new management of matings to avoid increasing of inbreeding rate.

Within the 84 bucks already genotyped, 11 (6 Alpines) were sires of the popQTL. Except these 11 bucks, 4 were sires of bucks from popIA born between 2007 and 2009. Six bucks from popIA had half sibs in popQTL and several parents of bucks from popQTL were grandparents of bucks from popIA. A reference population with a high number of effective founders like in this reference population is not ideal for genomic selection. Indeed GEBV accuracies were in inverse proportion to effective population size as shown Meuwissen’s (2009) prediction equation of GEBV accuracy.

III.2. LD extent in the population genotyped

III.2.A. LD within the population

Figure 13 shows LD extent in the multi-breed reference population evaluated with 1 946 females and 93 males genotyped. For marker distances lower than 0.5 Mb, LD extent was close to the one obtained for a population with constant effective population size of 1 000. However for larger marker distances LD extent in the reference population was close to the one calculated for population with effective population size of 250. Indeed LD extent at short distance between SNP reflects former selection because it was a function of effective population size many generations ago. LD at long distance reflects more recent history of the population. Indeed number of recombinations increases with time so LD which is result from past is LD at short distance between markers (Hayes, 2011).

The level of LD extent between two consecutives SNP (i.e. 0.06 Mb for goat chip) was 0.115. It was close to the LD extent found in Lacaune sheep (0.12) but lower to LD extent found in Holstein dairy cattle (0.3) (INRA internal communications). For small distance between markers LD extent in the multi-breed population was close to LD extent calculated separately in Alpine and Saanen breeds (Figure 14). However LD extent was lower for largest distance. Indeed Alpine and Saanen both originate from the Swiss Alps and widely crossbred were managed as purebred in France for more than forty years. Moreover LD in Saanen was slightly higher than LD in Alpine (0.135 vs 0.115 for marker distance of 0.06 Mb) whatever the marker distance considered. Level of LD extent in the goat reference population was small. It was not optimal for genomic selection (Calus et al, 2008). Nevertheless this level of LD was comparable to LD extent in Lacaune sheep for which genomic selection improves GEBV accuracies.
**Figure 15.** Persistence of LD extent between several breeds: Alpine and Saanen in the reference population if this study (a) and between ovine breeds (b)

**Figure 16.** Accuracies of GEBV function of heritability using three predicting equations from Goddard (2009), Deatwyler (2008) and Meuwissen (2009) with a reference population size of 705 or 1300 (705 males+ 1946 females i.e. 580 male equivalent).
III.2.B. Persistence of LD between Alpine and Saanen goats
Persistence of LD phases between Alpine and Saanen in the reference population (Figure 15 (a) in purple) for small distance between markers, 0.58 for 0.06 Mb was comparable to persistence of LD between two ovine French breeds Manech red faced and Manech black faced (Figure 15 (b) in pale orange). The two goat breeds (Alpine and Saanen) were genetically close for generations ago as the two Manech breeds are. However for more recent selection correlations of r values in the reference population (0.1 for 0.6 Mb) was closer to the one found between Lacaune and Manech black faced (0.09 for 0.6 Mb). The high level of LD phases persistence for consecutive marker distances, might improve accuracy of prediction by combining Alpine and Saanen animals in the same reference population.

III.3. Simulation of bucks genotypes
None parental incompatibility was detected in the simulations using Veriftyp2 software (Boichard and Druet, 2008) and LD extent for the 612 simulated genotypes conformed to the one in the reference population. Genotypes of animals with real genotypes were also simulated which allows verifying simulations. Correlations between allelic frequencies calculated on 85 real genotypes from popIA and the one calculated on simulated genotypes of the same bucks was quite high (84.4%). Average correlation between real genotypes of the 85 bucks and simulated genotypes of the same bucks was small (10.9%). Using the 685 simulated genotypes with the 2 039 real genotypes of the reference population did not appear to be feasible. In deed parental incompatibility in the mixed population (612 simulated genotypes + 2 039 real ones) were too high (36.7%). So to study the largest reference population feasible in goat species, all genotypes considered were simulated.

III.4. Prediction of GEBV accuracy level
III.4.A. In functions of the heritability and population size
Figure 16 shows the accuracy of GEBV for different values of heritability. Daetwyler et al (2010) and Goddard (2009)’s equations gave same results, and they agreed with Hayes et al ‘s study (Hayes et al, 2009b). Predictions of Meuwissen (2009) leaded to higher accuracies than the two others. They were close to real accuracies found in dairy cattle. Even with 1 300 genotypes available and a high heritability, GEBV accuracies for young males did not exceed 0.4. For milk yield (h²=0.3) accuracies of GEBV are expected to be 0.16 and 0.24 for fat percent (h²=0.4), it was really lower than those found in bovine. In the genotyped population (popIA and popQTL), and for milk yield, the CD for males was in average of 0.84 ± 0.38 and 0.55 ± 0.05 for females. For this trait, Meuwissen’s (2009) equation predicts GEBV accuracy
Figure 17. Accuracies of GEBV functions of reference population size (N) using three predicting equations from Goddard (2009), Deatwyler (2008) and Meuwissen (2009) with effective population size of 250 and a heritability of 0.3.

Figure 18. Real and predicted (by Meuwissen’s equation (2009)) GEBV accuracies for Lacaune sheep breed: Ne=230, L=23 M, and N= 2 089 or 1 024 rams (Baloche, personal communication)

Figure 19. Real and predicted (by Meuwissen’s equation (2009)) GEBV accuracies for French Holstein cows: Ne=36, L=30 M, and N= 4 000 bulls (Ducrocq, 2010).
for young males, comparable to CD calculated at birth (CD asc). In the genotyped population (males and females), ascendance accuracy was on average of 0.39 ± 0.02 (from 0.30 to 0.41) for milk yield. Predictions of GEBV accuracies were really lower than CD ascendance, genomic selection with such a reference population did not improve accuracy of genetic breeding values. Figure 17 shows GEBV accuracy predictions depending on the reference population size. To improve accuracy of genetic breeding value with genomic selection, a reference population of 4,500 buck genotypes will be necessary according to Meuwissen’s (2009) predictors. Nevertheless these predictions of GEBV accuracies do not take into account the high level of relationship within the reference population considered here. Therefore in this study accuracies were expected to be slightly higher than the ones predicted by those three equations. Besides Meuwissen’s equation (2009) seems to under predict GEBV accuracies. In Lacaune sheep breed (INRA internal communications), genomic evaluation was applied on two reference populations of 1,024 and of 2,089 rams. GEBV accuracies found for dairy traits were higher to those predicted by Meuwissen’s equation (2009) (Figure 18). For instance GEBV accuracy for milk yield was 0.37 (1,024 rams) whereas the equation predicted an accuracy of 0.16. Accuracies obtained in Holstein bulls were also higher than the one predicted. Indeed, genomic selection applied to 4,000 Holstein bulls (Ne=56 (Moureaux et al, 2000) and L=30 M) leads to GEBV accuracies between 0.60 and 0.70 for SCC (Ducrocq, 2010) whereas Meuwissen’s formula (2009) predicted an accuracy of 0.45 (Figure 19). Considering that, the predictions under predict real GEBV accuracies of 25 to 57%, milk yield should give a real GEBV accuracy between 0.20 and 0.25. It is already under CD ascendance for young males. Results of genomic selection in this study are not expected to be good.

**III.5.A. Function of LD extent**

Calus et al (2008) study relationship between accuracy of genomic breeding value and measure of linkage disequilibrium. Simulation of a bovine population with an effective population size (Ne) of 100, a trait with heritability of 0.5 or 0.1 and a reference population size of 1,100 bulls, allows them to plot GEBV accuracy obtained by Bayesian model as function of $r^2$ (Figure 20). In this figure juveniles were young animals with genotypes and without phenotypes, senior animals with genotypes and phenotypes. For juvenile animals, a $r^2$ level above 0.12 is necessary to noticed a significant increase of GEBV accuracies. Accounting for level of LD extent in the Alpine and Saanen reference population of this study, GEBV accuracies were expected to be around 0.7 for traits like fat and protein percent
Figure 20. Prediction of GEBV accuracies function of LD extent. Juvenile animals only genotyped, and senior animals phenotyped and genotyped.

Table 9. Correlations between solutions of genetic breeding values using GBLUP (GEBV) and solutions of genetic breeding values using BLUP (EBV), for different reference populations (see II.2.B.c).

<table>
<thead>
<tr>
<th></th>
<th>Case 20</th>
<th>Case 20+1,937</th>
<th>Case 93</th>
<th>Case 93 +1,937</th>
<th>Case 93</th>
<th>Case 93 +1,937</th>
<th>Case 705 si</th>
<th>Case 705 si +1,937 si</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>0.86</td>
<td>0.97</td>
<td>0.74</td>
<td>0.99</td>
<td>0.72</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>FY</td>
<td>0.97</td>
<td>0.99</td>
<td>0.69</td>
<td>0.72</td>
<td>0.68</td>
<td>0.76</td>
<td>0.92</td>
<td>0.99</td>
</tr>
<tr>
<td>PY</td>
<td>0.93</td>
<td>0.95</td>
<td>0.68</td>
<td>0.96</td>
<td>0.68</td>
<td>0.93</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>FP</td>
<td>0.87</td>
<td>0.91</td>
<td>0.66</td>
<td>0.74</td>
<td>0.62</td>
<td>0.76</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>PP</td>
<td>0.94</td>
<td>0.96</td>
<td>0.88</td>
<td>0.82</td>
<td>0.87</td>
<td>0.81</td>
<td>0.94</td>
<td>0.86</td>
</tr>
<tr>
<td>SCS</td>
<td>-</td>
<td>-</td>
<td>0.49</td>
<td>0.83</td>
<td>0.49</td>
<td>0.74</td>
<td>0.64</td>
<td>0.74</td>
</tr>
<tr>
<td>FU</td>
<td>-</td>
<td>-</td>
<td>0.82</td>
<td>0.84</td>
<td>0.82</td>
<td>0.81</td>
<td>0.90</td>
<td>0.81</td>
</tr>
<tr>
<td>UP</td>
<td>-</td>
<td>-</td>
<td>0.82</td>
<td>0.90</td>
<td>0.82</td>
<td>0.88</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>UFP</td>
<td>-</td>
<td>-</td>
<td>0.76</td>
<td>0.80</td>
<td>0.78</td>
<td>0.77</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>TO</td>
<td>-</td>
<td>-</td>
<td>0.86</td>
<td>0.87</td>
<td>0.85</td>
<td>0.84</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>RUA</td>
<td>-</td>
<td>-</td>
<td>0.87</td>
<td>0.75</td>
<td>0.87</td>
<td>0.72</td>
<td>0.90</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Figure 21. Genomic breeding values (GBLUP) function of genetic breeding values (BLUP), for SCS, and with a reference population of 93 males and females. Males in red and females in black considered are those phenotyped and genotyped.
(\(h^2=0.5\)) for young males. This prediction of GEBV accuracy was really higher than ascendance accuracy for fat and protein percent (0.43 in average). This prediction was necessarily over predicted because effective population size in goat reference population is 2.5 times the one considered in this graph. Nevertheless GEBV accuracy derived in this case seemed to be over predicted. In Lacaune sheep breed (INRA internal communication), with a \(r^2\) of 0.12, GEBV accuracy for protein percent (\(h^2=0.45\)) reached 0.39 for a reference population size of 1 024. The expected accuracy in this case was 0.7. Difference between expected accuracy from Calus et al (2008) and real one in Lacaune was too high to be induced by higher Ne, 270. Considering level of LD extent, effective and real reference population size, GEBV accuracies in the Alpine and Saanen reference population were not expected to be improved by genomic selection (Figure 20).

III.5. Results on genomic evaluation

Table 9 presents correlations between estimated breeding values using BLUP and the ones using GBLUP. Correlations ranged from 0.49 (SCS case 93) to 0.99 (FY case 705 + 1 946). In almost all cases, adding animals (males or females) in the reference population improved correlations (increasing of correlations from case 93 to case 705 + 1 946). For some type traits (FU, UFP, TO and RUA) and PP adding females in a reference population of 705 males decreased correlations between solutions (0.81 vs 0.90 for UFP). Correlations obtained for reference population with maximum of animals were quite high (from 0.74 to 0.99). The smallest correlations were obtained for SCS, from 0.49 to 0.83 followed by FP, from 0.62 to 0.81, and by UFP, from 0.76 to 0.90. Standard errors for milk yield FY, FP and type traits in genomic solutions were higher than those in BLUP solutions, from + 25% for milk yield (78.08 vs 69.09 for case 705 + females) to + 9% for UP (0.62 vs 0.55) (results not shown). These results showed that genomic information is different from pedigree one.

Figure 21 represents GBLUP solutions for SCS in function of BLUP solutions, for a reference population including 93 males and 1 946 females. Correlations between BLUP and GBLUP solutions are good for all traits, and for all animals except for 3 males without parents genotyped detected only in SCS. Their GBLUP solutions, from 0.2 to 0.8, were greatly higher than their BLUP solutions, from - 0.7 to - 0.3. Deviations between GBLUP and BLUP solutions for these 3 males were also high, in the 5% higher deviations, for two other traits: UFP and FP. Genetic correlations between UFP and SCS (- 0.24 in Alpine and - 0.19 in Saanen) and between FP and SCS (- 0.20 in Alpine and - 0.18 in Saanen) were the highest genetic correlations obtained between SCS and other traits (Clément et al, 2008). These
Figure 22. Average GEBV accuracy for animals genotyped and phenotyped (G+P) or young males (G) depending of population size in equivalent number of males: 673 including 1 946 females, i.e. 580 males equivalent and 93 males, or 705 males, or 1 285 (including 1 946 females i.e. 580 males equivalent, and 705 males).

Figure 23. Average GEBV accuracies (for 18 young males) obtained with a reference population of 705 males (adding or not females), and expressed in deviation to the average accuracies obtained with a reference population of 93 males (only males) or 93 males and 1 946 females (males + females).
results could indicate that these males are homozygote for a QTL concerning SCS. But further analyses are needed.

III.5.A. Effect of population size on GEBV accuracy

Figure 22 shows average GEBV accuracy for males with (G+P) or without (G) phenotyped depending of population size for PY, SCS and UP. Genomic accuracy calculated on males (G+P or G) were proportional to population size as for all traits. Increasing the population size from 93 to 705 or 680 improved accuracy of 30% for young males for PY. Adding 580 animals from a reference population with 705 males increases accuracy of only 18% for PY. Effect of population size on GEBV accuracies was higher for males genotyped and phenotyped than for young males, + 0.45 for former males and + 0.25 for young males, adding 1 192 males for UP. Increase of GEBV accuracy with population size was lower for SCS (trait with the lowest heritability) whatever the males considered than for other traits.

Figure 23 presents GEBV accuracy deviations for 18 young males between a reference population of 93 animals and another one adding 612 bucks for milk yield, FY, PP, SCS, UFP and TO. Results for protein yield were similar to those obtained for FY, and those obtained for FU and UP were similar to results for UFP. Increase population size of 612 did not improve GEBV accuracy more than + 0.31. Adding those animals to larger population (93 males and females compare to with only males) seemed to induce higher increase in GEBV accuracies on all traits, + 0.26 compare to + 0.24 for milk yield. Highest differences were noticed for type traits like UFP and RUA, around +19%, and the lowest one for fat percent, +8% (result not shown). But for FY, adding 612 animals on smaller reference population with only 93 males give better results, + 0.23 vs + 0.28. Increase of accuracy with population size was better for RUA, adding animals to a reference population of males and females and for FY considering a population with only males.

In Lacaune sheep breed adding 1 297 males in a population of 792 rams improved GEBV accuracy of 8% for milk yield (from 0.37 to 0.41), of 15% for FP and PP (from 0.39 to 0.45) and of 5% for SCS (from 0.37 to 0.39) (INRA internal communications). These results were close to those obtained in this study adding 1 946 females (580 males equivalent) to a reference population of 705 males for SCS, + 4%. But for milk yield, adding almost 600 individuals allowed a 22% increasing in GEBV accuracy, higher than the one found in Lacaune sheep breed.
Figure 24. Average deviation of GEBV accuracy for FP in reference populations with females from reference populations with only males: 20, or 93 (real genotypes), or 93 (simulated genotypes), or 705 males (simulated genotypes), for 18 young males and 75 males with daughters.

Figure 25. Deviation in GEBV accuracies for a reference population with 93 males and females (using simulated genotypes) from a reference population with 93 bucks, and with 705 and females males from 705 males.

Table 10. Average GEBV accuracies obtained for young males with a reference population of 705 males and 1,946 females compared to: predictions using equation Meuwissen’s (2009), results obtained in Lacaune sheep breed with a reference population of 792 rams, and to those obtained in Holstein dairy cattle with reference population of 3,242 bulls (Fritz et al, 2010).
III.5.B. Interest of female genotypes

Figure 24 shows GEBV accuracy deviations for FP with reference populations including 1 946 females and 20 or 93 or 705 males from reference populations without females. Increasing in accuracy noticed for young males, from 0.09 to 0.28 were higher than for males with daughters, from 0.01 to 0.14. Highest improvement of GEBV accuracy using female genotypes were obtained for the reference population of 93 males, + 0.28 with young males (using simulated genotypes) and + 0.14 with males phenotyped. Adding females in reference population of 705 males did not highly improved GEBV accuracies as it was expected, + 0.01 compare to + 0.12 in a reference population with only 93 bucks (using simulated genotypes). These results were similar for all other traits.

Figure 25 presents deviation of GEBV accuracy between reference population with females and without females for all traits considered. Increase of GEBV accuracy adding females in reference population did not exceed 0.2 for all traits except for PY, FP and PP (under 0.3). Highest improvement in GEBV accuracies was obtained for FP and PP, +0.28 i.e. + 1 650%, whereas lowest was noticed for FY, + 0.08 i.e. + 473 % considering a reference population of 93 males. GEBV accuracy were increased three times more adding females in a reference population of 93 males compare to one with 705 animals for all traits except SCS. Genomic predictions for SCS were increased fourteen time more, + 0.15 vs + 0.01 adding females with 93 males than adding them with 705 bucks. Adding females in reference populations to increase population size and so improve GEBV accuracies is a good solution for very small reference population size. It could also be interesting for larger population size but at a lower level depending on type of females genotyped, dams of sire or not. Benefits of female genotypes were higher for traits with high heritability like FP and PP. In this study, interest of genotyped females on GEBV accuracies compare to EBV ones, were lower than those find in Schaeffer’s study (2006), + 8.5% for 705 males and 1 946 females in reference population (result not shown), compare to + 25%. Nevertheless population size considered here is lower than those used in Schaeffer’s study. And female genotypes used were not the one of best females, only two dam of bucks on 1 946 females considered.

III.5.C. Interest of genomic selection

Average GEBV accuracy in a reference population of 2 642 animals (705 bucks + 1 946 goats) shown in Table 10 were up to 0.3 for milk yield, PY, FP, PP and RUA. Smallest GEBV accuracy (0.19) was obtained for SCS. These results were higher than those predicted by Meuwissen’s (2009) equation for all traits. But they were lower than those obtained for
Table 11. Deviation for protein yield between GEBV accuracy and ascendance accuracy for several reference populations studied, for young males.

<table>
<thead>
<tr>
<th>Case 93</th>
<th>Case 93 + 1946</th>
<th>Case 93 si</th>
<th>Case 93 + 1946 si</th>
<th>Case 705 si</th>
<th>Case 705 + 1946 si</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.84</td>
<td>-0.78</td>
<td>-0.70</td>
<td>-0.53</td>
<td>-0.14</td>
</tr>
<tr>
<td>Std</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Min</td>
<td>-0.89</td>
<td>-0.83</td>
<td>-0.75</td>
<td>-0.60</td>
<td>-0.35</td>
</tr>
<tr>
<td>Max</td>
<td>-0.77</td>
<td>-0.69</td>
<td>-0.65</td>
<td>-0.45</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Figure 26. Maximum deviation between GEBV accuracy and CD ascendance for young bucks, with a reference population of 2,642 individuals for all traits considered.

Table 12. CD ascendance and sire GEBV accuracy (CDg) for the young males (from 70 to 3 depending on the trait considered) with positive deviations between GEBV accuracy and CD ascendance.

<table>
<thead>
<tr>
<th>Milk yield, FY, PY</th>
<th>FP, PP</th>
<th>SCS</th>
<th>FU, UFP, TO</th>
<th>UP</th>
<th>RUA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD asc</td>
<td>0.40</td>
<td>0.40</td>
<td>0.29</td>
<td>0.32</td>
<td>0.25</td>
</tr>
<tr>
<td>Father CDg</td>
<td>0.96</td>
<td>0.95</td>
<td>0.57</td>
<td>0.83</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Lacaune sheep breed and for Holstein dairy cattle. Results for Holstein bulls considered a larger reference population than ours, and for in Lacaune sheep, only one breed compare to our multi-breed population, it could explain differences. Table 11 presents deviation between GEBV accuracy and CD ascendance for protein yield in several cases studied. Average deviations were improved by increasing population size with adding males or females in reference population. But all results were negative except the maximum of deviation for the biggest reference population, as for all dairy traits. Even a reference population of 2,642 animals did not allow improving accuracy of genetic breeding value at birth for all young males, - 0.08 in average. For type traits, a reference population with 705 males induced positive deviation for some individuals, + 0.16 and + 0.19 at maximum for UP and TO respectively (results not shown).

Maximum deviations between GEBV accuracy and CD ascendance are presented in Figure 26 for all traits considered here. Benefit of genomic selection in terms of accuracy of genetic breeding values at birth was higher for type traits, + 0.2 to + 0.25 at maximum compare to ascendance accuracy. The lowest deviations were obtained for fat yield and SCS, + 0.01 and + 0.04 respectively. Accuracy of genetic breeding values of some young bucks could be improved of 0.01 for FF and FP. These results are lower than those ones in Lacaune sheep breed with a reference population size of 792 rams: + 0.05 for milk yield, FY, PY and SCS, and + 0.06 on FP and PP (INRA internal communications). Nevertheless calculations of CD ascendance in sheep do not take into account ascendance of parents; it is not the case in goats.

Number of males with positive deviations between GEBV accuracy and CD ascendance were higher for TO, 70 bucks, than for other traits (34 in milk yield and PY). Few males had positive deviations for SCS and FY (4 and 3 respectively) where maximum of deviation was the less important. Males with positive deviations for milk yield had also positive deviations for other traits expect for FY for 31 bucks. Similarly males with positive deviations for RUA had positive deviations for all other type traits.

Table 12 presents GEBV accuracy of fathers and CD ascendance of these young males. Males with positive deviations for milk production traits were males with higher GEBV accuracy but also slightly higher CD ascendance, 0.40 in milk yield compare to 0.39 in average for the whole young males population. Their level of GEBV accuracy could be explained by the fact that their sires were genotyped, with the consequence also for the fathers to have slightly higher GEBV accuracies than the ones find for all sires, 0.95 compare to 0.94 for FP and PP. For type traits and SCS, males with positive deviations were males with lower
Figure 27. GEBV accuracy of the twenty with daughters depending on number of genotyped daughters
CD ascendance and with average GEBV accuracy. Some of these bucks have fathers genotyped but not all. The CDs ascendance calculated for those traits were lower than those calculated for milk production traits. These traits were recorded more recently than other traits, so CDs ascendance for parents were lower.

Figure 27 shows GEBV accuracy of the twenty males from popQTL obtained with a reference population of 2,642 individuals in function of number of genotyped daughters. These results are similar to those obtained for each trait. It seemed to be no relation between the numbers of genotyped daughters and the GEBV accuracy obtained. But the difference of number of daughters between those males is not high and did not allow noticing a relation. Nevertheless we have observed that males with higher GEBV accuracy were those with genotyped sires.

Considering a multi-breed reference population of 705 males and 1,946 goats, genomic selection did not appear to improve accuracy of genetic breeding value at birth. Nevertheless these results are dependent on the way of accuracy calculation. But some characteristics of ideal reference population in goat specie have been pointed out. Increasing population size would be the better solution to improve genomic prediction with females or with males highly related with those already genotypes like sires.

Conclusion

This first study on genomic selection in goat species allowed characterizing French reference population available at this time. Considering this multi-breed population, the number of effective founders, 250, was high, and level of inbreeding, 2.29% in the whole population was weak. These results indicated that this population is not ideal for genomic selection. But high level of relationship within the population, sires, grand-sires and cousins could compensate this problem.

Persistence of LD phases between Alpine and Saannen breeds was high, 0.6, for short marker distance, 0.06 Mb. It justified including the two breeds together in the same reference population. Nevertheless level of LD extent for average distance between two markers, 0.06 Mb was 0.115, lower compare to Holstein dairy cattle LD level. A reference population with such a LD extent was not perfect for genomic selection.

Interest of genomic selection is to improve GEBV accuracies of animals at birth. Genomic accuracy predicted by Meuwissen’s equation (2009), which is under estimated, was 58 % under the average CD ascendance using the highest reference population available of
705 males and 1,946 females. In this context a population of 4,500 males seemed to be necessary to reach ascendance accuracy level. Genomic selection was not expected to improve GEBV accuracy at birth. As it was expected increase population size induced improvement in GEBV accuracy for all animals. Adding 612 animals to a population of 93 males increased GEBV accuracies in average by 12% and 5% respectively for protein yield and somatic cell score, than adding 580 males equivalent from a reference population of 705 males. Increasing GEBV accuracy with the reference population size was better for fat yield, udder floor position and rear udder attachment. Using females to increase population size was more efficient for small population size, + 0.28 with 93 males compare to + 0.18 with 705 males for fat percent.

Accuracies of genetic breeding values at birth were not improved with genomic selection even of the largest reference population considered. However a maximum increase in accuracy of 0.25 on rear under attachment was noticed for some males with genotyped sires. These results were conditioned by level of CD ascendance, really higher of former recorded traits like milk production ones. They were lower than those obtained in Lacaune sheep partly because CD ascendance is under estimated in dairy sheep.

Results on GEBV accuracy could be improved considering the 768 additional genotypes available at the end of the year. Best females like dams of bucks could be genotyped to increase population size. Moreover considering the two goat breed together might not improve GEBV accuracy as well as it is expected. In sheep, a multi-breed (Manech red faced and Manech black faced) reference population did not improve GEBV accuracies of both breeds compare to consider each of us separately; even if Manech black faced had a really small reference population (300 animals (Beltran De Heredia et al, 2011)).

However, this study on genomic selection allowed us to suspect QTL for somatic cell score trait, needing other analysis of QTL detection. With this study we have pointed out difficulties to consider reference population in goat species. However we have found leads of improvement to consider genomic selection feasible in goats like increasing number of animals in reference population with dams of bucks for instance or reviewing accuracies calculations.
Acknowledgements

This study was funded by the French Genovicap and Phenofinlait programmes (ANR, Apis-Gène, CASDAR, FranceAgriMer, France Génétique Elevage and the ministry of agriculture) and the European 3SR project. I thank all the people responsible for goat selection in INRA SAGA: Virginie Clément, Hélène Larroque, Isabelle Palhière and Rachel Rupp for their help to overcome problems with dataset and their comment during this internship. I thank also Christèle Robert-Granié for her precious advices and help on the implementation of genomic selection and for all statistical problems encountered. I would like to address thank to Gwenola Tosser-Klopp from LGC lab for her help with genotyping data set. For her availability, and our discussion about thesis and goat genetic, I want to thank Cyrielle Maroteau. And for his help with Fortran programs of LD calculations and with awk programming for sorting genotypes, I would like to thank Guillaume Baloche. I want also to thank Julien Jacquin for his allele frequencies calculation program and for his help for understanding Fortran programming. Finally I want to thank all the people from SAGA lab for their kindness.
References
- Sørensen A. C. , Sørensen M. K. (2010) Genotyping both males and females is favorable in genomic dairy cattle breeding schemes, Proceedings of 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, August 1-6, 2010 Liepzig, Germany.
- Sullivan P. (2010) CrEDC_5e.c, Compute EDCs for submission to Interbull for any number of traits, version August 2010.


### Annex A

Synthesis of the article dealing with the multi breed or admixed reference population

<table>
<thead>
<tr>
<th>Publication</th>
<th>Specie and breed(s)</th>
<th>Model(s)</th>
<th>Variable studied</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brøndum et al (2011)</td>
<td>Dairy cattle Danish, Swedish and Finnish</td>
<td>Bayes</td>
<td>DGV reliability</td>
<td>Three breed reference population increase reliability of 0.03 in Danish and 0.07 in Swedish and Finnish compare to national reference</td>
</tr>
<tr>
<td>De Roos et al (2009)</td>
<td>Dairy cattle (simulated bulls)</td>
<td>Bayes</td>
<td>GEBV reliability</td>
<td>Adding animals from population B to A increase reliability in animals of population B but not in population A.</td>
</tr>
<tr>
<td>Ibáñez-Escriche et al (2009)</td>
<td>Beef cattle (simulated)</td>
<td>Bayes B</td>
<td>GEBV accuracy</td>
<td>Accuracy decrease with the distance between breed. Adding across-breed effect to estimate SNP effects does not improve accuracies</td>
</tr>
<tr>
<td>Kizilkaya et al (2009)</td>
<td>Beef cattle Angus and others</td>
<td>BayesB</td>
<td>Correlations predicted - simulated EBVs</td>
<td>Correlations between were better using purebred (PB) than multi-breed (MB) reference population and for Angus (in PB and MB populations)</td>
</tr>
<tr>
<td>Lund et al (2010, 2011)</td>
<td>Dairy cattle Holstein (4 000 bulls)</td>
<td>Bayes</td>
<td>DGV reliability</td>
<td>Using across country reference population gives higher DGV reliability than only one country (+ 32% for the German, + 20% for others)</td>
</tr>
<tr>
<td>Makgahlela et al (2010)</td>
<td>Dairy cattle, Danish, Swedish and Finnish</td>
<td>GBLUP</td>
<td>DGV reliability</td>
<td>Reliability with or without the specific breed effect are similar, except for protein yield where breed effects is better (0.31 against 0.28).</td>
</tr>
<tr>
<td>Pryce et al (2011)</td>
<td>Dairy cattle, Holstein, Fleckvieh and Jersey</td>
<td>GBLUP BayesA</td>
<td>GEBV accuracy</td>
<td>Adding breeds in the reference population to predict one breed is not efficient but to predict three breeds it is better.</td>
</tr>
<tr>
<td>Toots et al (2009)</td>
<td>Beef cattle (simulated)</td>
<td>BayesB</td>
<td>GEBV Accuracy</td>
<td>Using an admixed training population leads to same accuracy than using purebred population</td>
</tr>
<tr>
<td>Varona et al (2010)</td>
<td>Beef cattle simulated</td>
<td>Bayes</td>
<td>GEBV accuracy</td>
<td>If population highly related, same accuracy obtained considering same effects of SNP across breeds or different but correlated effects.</td>
</tr>
</tbody>
</table>
### Annexe B

**Synthesis of article dealing with the integration of female genotypes in the reference population**

<table>
<thead>
<tr>
<th>Publication</th>
<th>Type of female used</th>
<th>Species and breed</th>
<th>Trait</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buch <em>et al</em> (2010)</td>
<td>2000 best female on PA (parent average)</td>
<td>Dairy cattle</td>
<td>Protein yield (h²=0.3) mastitis resistance (0.04)</td>
<td>Adding females increase genetic gain of 21% compare to progeny testing and +18% compare to only males</td>
</tr>
<tr>
<td>Calus <em>et al</em> (2010)</td>
<td>2000 commercial cows + dams of bulls</td>
<td>Dairy cattle</td>
<td>New trait h²=0.3</td>
<td>Adding 1000 females from reference population with 1000 females increase reliability of 40%</td>
</tr>
<tr>
<td>Jiménez <em>et al</em> (2010)</td>
<td>Commercial cows + dams of bulls</td>
<td>Dairy cattle</td>
<td>Two simulated traits (h²=0.25 or 0.1)</td>
<td>Adding 8% of females in a reference population with 2% of females increase accuracy of 4%, using top female compare to random one (+6%)</td>
</tr>
<tr>
<td>Mc Hugh <em>et al</em> (2011)</td>
<td>2625 commercials cows</td>
<td>Dairy cattle</td>
<td>Simulated trait h²=0.3</td>
<td>Using males and females reference population increase accuracy of 23% compare to progeny testing, and genetic gain of +0.17 σₐ (+100%).</td>
</tr>
<tr>
<td>Pryce <em>et al</em> (2010)</td>
<td>Selection on dams of bulls using MOET</td>
<td>Dairy cattle</td>
<td>Simulated trait</td>
<td>Intensive selection on females increase genetic gain of +198% compare to progeny testing</td>
</tr>
<tr>
<td>Schaeffer (2006)</td>
<td>2000 dams of bulls</td>
<td>Dairy cattle</td>
<td>Simulated trait</td>
<td>Adding females in reference population increase GEBV accuracies of 23% compare to progeny testing.</td>
</tr>
<tr>
<td>Schrooten <em>et al</em> (2005)</td>
<td>Dams of bulls</td>
<td>Dairy cattle</td>
<td>Total merit index</td>
<td>Using only females in reference population decrease genetic gain if variance proportion explained by QTL is under 25%, otherwise it increase of 18.8%</td>
</tr>
<tr>
<td>Sørensen and Sørensen (2010)</td>
<td>Several proportion of best females</td>
<td>Dairy cattle</td>
<td>total merit index (production + functional traits)</td>
<td>Adding 25% of females in reference population increase genetic gain of 26% from 0 to 25%, but decrease of 4% from 60 to 85% of females.</td>
</tr>
<tr>
<td>Spangler <em>et al</em> (2008)</td>
<td>Commercial females or females selected</td>
<td>Dairy cattle</td>
<td></td>
<td>Adding females increase APTG of 3% and 7.7% respectively with unselected and selected females compare to only males.</td>
</tr>
<tr>
<td>Verbyla <em>et al</em> (2010)</td>
<td>527 heifers</td>
<td>Dairy cattle</td>
<td>Energy balance</td>
<td>Reliabilities are increase of 39% using males and females in reference population compare to progeny testing.</td>
</tr>
</tbody>
</table>

**APTG**: Average probability for each animal that his true genotype is identified
Annexe C:
Models used in genetic evaluation for achievement of phenotypes

**Model for official genetic evaluation YD\textsubscript{allfem} and DYD\textsubscript{allfem}**

4 Regions: North-East (1), South-West (2), North-East (3), South-East (4)
Lactation number: L1, 2, 3, 4, 5, 6 and +
Fixed effects on the mean:
- Farm*Year*Lactation number (1, 2, or 3and +)
- Age at kidding*Year*Region
- Month of kidding*Year*Region
- Last of dry period*Year*Region
Random effects on the mean:
- Permanent environment (several lactations for a goat)
- Genetic value: Group of unknown parents (Saanen, Alpine, Crossbred):

Effects on residual variance:
- Year*Region*Lactation number (fixed)
- Farm*Year*Lactation number (random)

**Model genetic evaluation on females genotyped YD\textsubscript{femg} and DYD\textsubscript{femg}**

2 Regions: North (1 + 3), South-West (2)
Lactation number: L1, 2 and +
Fixed effects on the mean:
- Farm
- Age at kidding*Region
- Month of kidding*Region
- Last of dry period *Region
Random effects on the mean:
- Permanent environment (several lactations for a goat)
- Genetic value: Group of unknown parents: