# Assessment of a genomic design for a French meat sheep breeding program

*J. Raoul[1,2], A.A. Swan[3] & J.M. Elsen[2]*

*[1] Institut de l'Elevage, BP 42118, 31321 Castanet-Tolosan, France*
*[2] GenPhySE, INRA, 31326 Castanet-Tolosan, France*
*[3] Animal Genetics and Breeding Unit, University of New England, 2350 Armidale, Australia*

## Summary

The majority of French meat sheep populations use both insemination (AI) and natural mating sires. Usually, AI sires are progeny tested and then the best are used as proven sires to produce male candidates. A breeding program based on genomic selection would be an alternative. Using a stochastic model, where both individuals and their genomes were simulated, we assessed a genomic breeding program design. The reference population was based on sires genotyped with a medium density panel (MD = 50K SNPs), including two to ten generations of sires born before the implementation of the genomic scheme and all sires born thereafter. For sire replacement, newborn progeny were first preselected on parent average genomic estimated breeding values (GEBV) and then genotyped with a very low density panel (VLD = 1K SNPs). MD genotypes of candidates were imputed using the software Fimpute and GEBV computed with a single step Genomic BLUP animal model using the software Blupf90. Males selected for replacement were then genotyped with the MD panel to update the reference population. We assessed the sensitivity of genetic gain to various sizes of the initial reference population and compared the genetic gain of genomic and classical breeding program designs at a fixed total cost with three different price levels for VLD genotyping. Within the range of values assessed for the size of the initial reference population and the price level of VLD genotyping, no significant differences between genomic breeding schemes was observed. At a fixed total cost, the annual genetic gain was higher for genomic designs (+18%) than for the classical design.

*Keywords: breeding program, stochastic simulation, genomic selection, imputation, sheep*

## Introduction

Implementing a genomic breeding program is still a challenge for small meat sheep populations, where estimated breeding values (EBVs) of both artificial insemination (AI) and naturally mated sires have limited accuracy due to low progeny numbers per sire, compared to dairy cattle for example. Reaching a high genomic prediction accuracy, such as predicted by Daetwyler *et al.* (2008), would require establishment of a reference population based on a large number of animals, which is infeasible for small sheep breeding nuclei. Including records and genotypes from lower tiers of the population can solve this issue (Santos et al., 2017), although this might be difficult to achieve due to the lack of recording in commercial flocks. However, when the genotyped reference population included all sires and grandsires of candidates, Raoul *et. al* (2017) found an additional genetic gain of + 21.6% compared to a classical scheme when a full genomic scheme was adopted, including steps to impute 50K

SNPs “Medium Density” (MD) genotypes from 1K “Very Low Density” (VLD) SNP genotypes of male candidates.

French meat sheep breeding programs are based on collective management of males that is coordinated by breeding societies. Male candidates enter into phenotype recording stations where AI and natural mating sires are selected. AI males are progeny tested and at least two years later, based on their progeny records, the best AI sires are used as proven sires to produce male candidates. In a genomic design, male candidates might first be selected on parent average Genomic EBV (GEBV) and genotyped at MD after imputation from a VLD panel. Then, replacements might be selected among candidates on their own GEBV. Male candidate genotypes would be purchased by the breeding society. To implement such a design, the cost needs to be similar to the cost of the current breeding design, because the profitability at the nucleus level remains a critical factor in sheep (van der Werf and Banks, 2017). The current investment dedicated to progeny testing would be allocated to VLD genotyping of male candidates and MD genotyping of sires (initial reference population and newly selected sires).

Focusing on a breeding program for a small population of purebred sheep, we used stochastic simulation to assess (1) the sensitivity of the additional gain from genomic selection to lower sizes of the initial reference population, and (2) the genetic gain of genomic and classical designs at a fixed total cost for three price levels of VLD genotyping.

## Materials and methods

Using a stochastic model, we simulated individuals and their genome based on 50 K real genotypes. To obtain a founder population of 5000 females with a pedigree structure, we first realized random reproductive cycles. Then we applied a classical design based on the progeny testing of AI sires over 10 years. Only one maternal selected trait was considered. The next fifteen years, we applied either a classical or a genomic breeding program design. The model included the establishment of the founder population, the simulation of QTL and phenotypes, the simulation of genotypes based on the VLD and MD SNP panels, the imputation of MD genotypes using FImpute software developed by Sargolzaei et al. (2014), and the estimation of breeding values by the single step GBLUP using the Blupf90 software developed by Misztal et al. (1999). The key design information relevant to this study is described below, and the model is fully described in Raoul et al. (2017).

Around 5000 females divided in 10 flocks were recorded (one record per year) for the selected trait. Each year, half of the breeding females were selected on EBV and mated to an AI sire. Females that did not conceive to AI were mated to a natural mating sire, along with females that were specifically selected for natural mating. The number of progeny per dam depended on the mode of reproduction and parity. Some dams were randomly culled after a reproductive cycle and the maximum parity was seven. Around 24% of dams were replaced per year by females that were preferentially chosen among newborn progeny from AI matings. No selection trait was considered for newborn females before they were mated. There was no difference between the classical and genomic breeding program designs regarding female replacement.

Male candidates were preselected among newborn progeny on their parent average EBV in the classical design and on their parent average GEBV in the genomic design. In practice, all candidates were born from proven AI sires (classical design) and from genomic AI sires (genomic design), given their genetic superiority. In the classical design, the ten candidates with the highest parent average EBV were selected and mated across flocks by AI to be progeny tested. Two years later, the five AI sires with the highest EBV (including progeny records) were selected as proven AI sires and used at most for four years. In the

genomic design, male candidates were genotyped with the VLD panel and their MD genotypes were imputed. The ten candidates with the highest GEBV were selected as genomic AI males and used at most for two years. Naturally mated male replacements were selected among candidates that not selected for AI, based on their parent average EBV in the classical design and on their GEBV in the genomic design. Naturally mated males were used at most for four years during which no further selection was applied. In the genomic scenario, all selected sires were then genotyped with the MD panel to update the reference population.

The annual genetic gain was estimated as the slope of the regression of the average true breeding value (TBV) of first parity females on time for years 10 to 15 and 10 to 25. The means and standard deviations presented are based on 50 replicates.

To assess the sensitivity of the genetic gain to the size of the initial reference population, we built different initial reference populations: in addition to all sires in use when the design shifted to genomic selection (year 10), were included other sires that were used in the 2, 4, 6, 8 or 10 years prior to the start of genomic selection. Regardless of the size of the initial population, all sires selected in subsequent years were included in the reference population and the number of genotyped candidates per year was constant (270). To compare the classical and genomic designs at a fixed total cost, we first determined the variable costs in the genomic and classical scheme: MD and VLD genotyping and the costs of maintaining AI sires. The money saved by using AI sires for a shorter time period in the genomic design was invested in MD genotypes (75€) of the initial reference population, in selected sires in subsequent years, and in VLD genotypes of male candidates. First, we determined the optimal investment in either an increase in the initial reference population size, or an increase in the number of candidates per year over a fifteen year time period, for three price levels of VLD genotyping. Then, the genetic gain of the classical and optimized genomic designs (time 10-25) were compared for the three price levels of VLD genotyping.

## Results and Discussion

Table 1 gives the annual genetic gain for the classical and genomic designs, with sizes of the initial reference population. We observed a slight sensitivity of the genetic gain from 10 to 15 years to the size of the initial reference population but the gain from 10 to 25 years did not differ significantly. These results indicate that the presence of candidate sires in the ongoing reference population is more important than the size of the initial reference population for the designs we evaluated. The average accuracy of the GEBV of selection candidates in the genomic designs was 0.12 points higher than the average accuracy of EBV in the classical scheme (0.39) (results not shown).

*Table 1. Annual Genetic gain for the classical and genomic designs, with different sizes of the initial reference population (50 replicates)*

| Design | | CS[1] | GS[2] | | | | |
|---|---|---|---|---|---|---|---|
| Initial reference population size[3] | | | 175 | 210 | 300 | 400 | 500 |
| Years prior to start of GS[4] | | | 2 | 4 | 6 | 8 | 10 |
| Gain (genetic STD/year) | | | | | | | |
| 10-15 years | mean | 0.173 | 0.173 | 0.175 | 0.179 | 0.185 | 0.188 |
| | STD | 0.021 | 0.021 | 0.023 | 0.020 | 0.020 | 0.022 |
| 10-25 years | mean | 0.167 | 0.193 | 0.192 | 0.196 | 0.197 | 0.197 |
| | STD | 0.014 | 0.018 | 0.019 | 0.019 | 0.017 | 0.018 |
| Gain, 10-25 years (%, classical =100) | | 100.0 | 115.4 | 114.8 | 117.2 | 118.0 | 118.4 |

[1]Classical design, no genomic information was used. Male replacements were selected on progeny records.

[2]Genomic design, male replacements were selected on their GEBV among 270 genotyped candidates.
[3]The initial reference population included sires used at time 10 and other sires used from previous years (from the last two years to the last ten years)
[4]Number of years prior to the start of genomic selection for inclusion of sires in the reference population

In Raoul *et al.* (2017), a reference population of only sires was sufficient to impute MD genotypes from VLD genotypes with a concordance rate close to 0.95, although a scenario where the imputation was performed without pedigree information gave substantial losses in genetic gain (-22.2%). Previous studies (e.g. Clark et al. 2012) also highlighted the importance of close relatives in the reference population for the accuracy of genomic prediction. Misztal et al. (2013) reported that using single step GBLUP increased accuracy compared to multi-step GBLUP. The use of a single step GBLUP would increase the efficiency of genomic breeding programs especially when the number of genotyped animals is limited. Given the heritability of the trait (0.25), it is highly likely that the moderate accuracy of the GEBV might be enhanced by a substantially larger reference population. However, although the increase in accuracy of the GEBV of selection candidates was moderate, the genomic design was still superior to the classical design based on progeny testing.

Table 2 gives the maximal annual genetic gain for a genomic design at a fixed total costs for three price levels of VLD genotyping and price of 75 € for MD genotyping. The optimized combination of size of the historical reference population and the number of genotyped candidates per year is also given.

*Table 2. Maximal annual genetic gain for a genomic design at a fixed total costs based on three price levels for VLD genotyping.*

| | VLD genotyping cost (€/unit) | | |
|---|---|---|---|
| | 7.5 | 15 | 22.5 |
| Gain (genetic standard deviations) | 0.197 | 0.194 | 0.193 |
| Initial reference population size: years prior to start of GS[1] | Y=10 | Y=8 | Y=4 |
| Number of VLD genotypes per year[2] | 270 | 200 | 215 |

[1]Number of years prior to the start of genomic selection for inclusion of sires in the reference population
[2]VLD: very low density SNPs panel on male candidates

Regardless of the VLD genotyping cost, the genomic design was superior to the classical design (0.167). Surprisingly, no significant differences in genetic gain were observed for the three VLD genotyping cost levels. A reduction in the historical reference population size and a reduction in the number of candidates had little impact on genetic gain. In this study, for practical reasons, the maximum number of preselected males genotyped was 300 and sires were selected among 270 genotyped candidates (10% were culled or died for other reasons). The selection intensity of genomic sires was little affected by reducing the number from 270 (intensity = 2.186) to 200 (intensity = 2.063). This slight reduction in intensity and the low sensitivity to the initial reference population size may explain the stability of results across the range of genotyping costs considered.

We conclude that, for small sheep populations, a genomic design based on a reference population that includes all sires and VLD genotyping of male candidates is a viable alternative to current breeding programs based on progeny testing.

## List of References

Clark SA, JM Hickey, HD Daetwyler, JH van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. GSE. 44:4.

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:E3395.

Misztal I. 1990. Complex models, more data: simpler programming. Interbull Bull. 20:33-42.

Misztal I, SE Aggrey, WM Muir. 2013. Experiences with a single-step genome evaluation. Poultry Sci. 92:2530–4.

Raoul J, AA Swan, JM Elsen. 2017. Using a very low density SNP panel for genomic selection in a breeding program for sheep. GSE. 49(1):76

Santos BFS, JHJ van der Werf, JP Gibson, TJ Byrne, PR Amer. 2017. Genetic and economic benefits of selection based on performance recording and genotyping in lower tiers of multi-tiered sheep breeding schemes. GSE.;49:10.

Sargolzaei M, JP Chesnais, FS Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.

van der Werf J, R. Banks. 2017. Effective breeding programmes down under. JABG. 134:1–2.