Feasibility of genomic predictions of Sarda breed rams using a female reference population

M.G. Usai¹, S. Salaris¹, S. Casu¹, T. Sechi¹, S. Miari¹, P. Carta¹, & A. Carta¹

¹ AGRIS SARDEGNA, Servizio Ricerca per la zootecnia, località Bonassai – 07100 Sassari, Italy
gmusai@agrisricerca.it (Corresponding Author)

Summary

In this study we performed a first assessment of the utility of genomic predictions for milk yield in Sarda sheep. The reference population consisted of 3,731 ewes from a nucleus flock. The validation population consisted of 537 rams from the herd book. Breeding values of rams were estimated by using performances from the nucleus and either pedigree or genomic information and the accuracy of estimated breeding values compared. A measure of the information on relatives in the reference sample was calculated from the pedigree-based relationship matrix. Results showed that the use of genomic information improves the accuracy of estimated breeding values relative to using standard pedigree. The proposed measure of the information on relatives was a good predictor of the accuracy of genomic predictions. Results were discussed to identify relationships of the accuracy of genomic predictions with information on relatives and category of rams. We conclude that genomic predictions of Sarda rams based on a female reference population can be used for selective breeding.

Keywords: dairy sheep, genomic prediction, female reference population

Introduction

The use of genomic predictions for selective breeding in dairy sheep must be adapted to the specific pre-existing breeding schemes taking costs and benefits into account. The dairy cattle approach (Boichard et al., 2015) has been successfully applied in some dairy sheep breeds, such as the French Lacaune, where the large use of AI and the accurate recording schemes for milk yield and composition and for functional and health traits, have made genomic selection of breeding males profitable (Baloche et al., 2014). In other breeds, such as the Sarda, the lack of accurate recording schemes for traits other than yields and the difficulty to create a sire reference population due to the low usage of AI, make this approach not feasible (Carta et al., 2009). In these circumstances, Sarda breeders are assessing the possibility of using a flock of ewes, originally established for QTL detection on a large number of traits, as a female reference population to underpin genomic predictions for breeding rams to be used in the herd book (HB). In this study, we performed a first assessment of the accuracy of genomic predictions for milk yield, with the aim of identifying useful criteria to optimize the size and the structure of the female reference population and modulate the flow of breeding animals from and toward the HB.

Materials and methods
Female reference population

The Sarda female reference population (FRP) consists of approximately 1,000 milking ewes with an annual replacement of about 25%. Replacements are generated by mating adult ewes of FRP with rams coming the Sarda HB. Ewes of FRP are raised in an experimental farm that has a typical Sardinian dairy sheep farming system. The original aim of the FRP was to detect QTL segregating in the Sarda breed. Thus, the number of daughters per ram was 40 ewes on average until 2009. It was reduced to 9 after 2010 with the aim of increasing the number of breeding rams with daughters in FRP and consequently the number of represented bloodlines from HB. So far, 3,949 ewes have been generated by 161 rams. The FRP ewes are routinely recorded for several functional, health and production traits. In the present study, we focus on milk yield, which is quite accurately recorded also in HB. In the end, the reference sample consisted of 3,731 FRP ewes born from 2000 to 2015 for which SNP genotypes and 13,059 lactation records were available.

Male validation population

The validation sample consisted of 537 HB rams. In the Sarda HB the progeny test of rams is carried out either by AI or by controlled natural mating (CNM). Annually, breeding values for milk yield (EBV) are estimated by a classical BLUP-animal model using a pedigree-based relationship matrix and lactation records from the HB. Records of FRP ewes are not considered. Rams of the validation sample were chosen to represent all categories of breeding males: 144 rams involved in the AI program with daughters in FRP, 150 rams involved in the AI program without daughters in FRP and 288 rams used by CNM in HB with no daughters in FRP. For all these rams official EBV for milk yield were available.

Genotypes

Both reference and validation samples were genotyped with the Illumina Inc. OvineSNP50 Beadchip. SNP editing was performed on the 26 OAR autosomes using call rate and MAF thresholds of 95% and 1% respectively. After quality control, 43,390 SNP were retained.

Genetic evaluation

FRP lactation records were used to estimate breeding values of the validation sample by either pedigree or genomic information in order to compare their accuracies and verify the consistency with the official EBV calculated using lactation records from the HB. The numerator relationship matrix (A) was formed using all HB animals that had relatives both in the reference and in the validation samples. In total, 11,882 individuals were retained. The genomic relationship matrix (G) was formed using genotypes of the reference and validation samples (Hayes et al., 2009). The following repeatability animal model was fitted by ASREML (Gilmour et al., 2009) for both pedigree (PBLUP) and genomic (GBLUP) predictions:

\[ y = Xb + Zu + Wp + e \]

where \( y \) is the vector of the observations (13,059 lactation records); \( b \) is the vector of fixed effects; \( u \) is the vector of animal breeding values; \( p \) is the vector of random permanent environmental
effects; \( \mathbf{e} \) is the vector of random residuals and \( \mathbf{X}, \mathbf{Z} \) and \( \mathbf{W} \) are incidence matrices relating records to fixed effects, breeding values and permanent environmental effects respectively. Fixed effects included in both models were: year-management-group, year-month of lambing-parity-age class and milking length within age class (yearlings or adult) as covariate. In both models, vectors \( \mathbf{p} \) and \( \mathbf{e} \) were assumed distributed as \( \mathbf{p} \sim \mathcal{N}(0, \mathbf{I}\sigma_p^2) \) and \( \mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2) \), where \( \sigma_p^2 \) and \( \sigma_e^2 \) are the **permanent environmental and residual variance** respectively. In PBLUP, \( \mathbf{u} \) was the vector of additive genetic values (PBV), assumed distributed as \( \text{PBV} \sim \mathcal{N}(0, \mathbf{A}\sigma_a^2) \), where \( \mathbf{A} \) is a numerator relationship matrix and \( \sigma_a^2 \) is the additive genetic variance. In GBLUP, \( \mathbf{u} \) was the vector of genomic breeding values (GBV) assumed distributed as \( \text{GBV} \sim \mathcal{N}(0, \mathbf{G}\sigma_g^2) \), where \( \mathbf{G} \) is a genomic relationship matrix and \( \sigma_g^2 \) is the genetic variance.

**Accuracy of predictions and effect of information on relatives**

The accuracy of PBV of the \( i \)-th ram from the validation sample was calculated as \( r_{PBVi} = \{1 - [\text{SE}_{i}^2/(\mathbf{A}_{ii}\sigma_a^2)]\}^{1/2} \), where \( \text{SE}_{i} \) is the standard error of the PBV prediction and \( \mathbf{A}_{ii} \) is the diagonal element of \( \mathbf{A} \) for ram \( i \). Similarly, the accuracy of the GBV predictions of the \( i \)-th ram was calculated as \( r_{GBVi} = \{1 - [\text{SE}_{i}^2/(\mathbf{G}_{ii}\sigma_g^2)]\}^{1/2} \), where \( \text{SE}_{i} \) is the standard error of the GBV prediction and \( \mathbf{G}_{ii} \) is the diagonal element of \( \mathbf{G} \) for ram \( i \). Correlations between PBV and GBV predictions and with the corresponding official EBV were also calculated. In order to measure for each ram the amount of information on relatives in FPR, the section of the \( \mathbf{A} \) matrix that included all FRP ewes and the ram itself was inverted and the diagonal element corresponding to the ram (diA) was retained. **We used diA instead of the corresponding element from \( \mathbf{G} \) since \( \mathbf{A} \) is the only information available before genotyping.** The ability of diA to predict \( r_{GBV} \) was investigated by fitting a specific function.

**Results and discussion**

**Accuracy of genomic evaluations**

The heritability estimated by PBLUP was higher than that estimated by GBLUP (0.41 vs 0.35). This result is consistent with other studies, both in sheep and cattle, but there is no clear explanation for it. Repeatability estimates were similar (0.59 vs 0.58) for the two models. These estimates indicate that the use of \( \mathbf{A} \) or \( \mathbf{G} \) affects the allocation of the individual variance to genetic or permanent environmental components. The correlation between predictions of GBV and PBV in the validation sample was 0.92 suggesting that the sample of selected top rams may be re-ranked using GBLUP instead of PBLUP. Moreover, \( r_{GBV} \) was 31% higher than \( r_{PBV} \) and the correlation of GBV with official EBV was 71% higher for predictions of GBV than for predictions of PBV (Table 1). These results demonstrate that the use of genomic information improves the reliability of breeding values compared to using pedigree.

**Effect of information on relatives on accuracy of genomic evaluation**

As expected, diA is a good predictor of \( r_{GBV} \) (Figure 1). This parameter could be used to select the new sires of FRP in order to maximize the number of HB rams with a sufficient \( r_{GBV} \) and, thus, the number of rams that will be accurately evaluated by GBLUP after genotyping. In order to explore the pattern of diA and \( r_{GBV} \) according to the category of rams (AI or CNM rams), sampled rams were
divided into three groups based on the level of diA: highly related (diA > 2.5), medium related (diA > 1.25 and < 2.5) and lowly related (diA < 1.25). The group with the highest diA was composed of rams involved in the AI program, the most having daughters in FRP (20 on average). Even the group with medium diA was mainly composed of rams involved in the AI program. However, a few of them had daughters in FRP (2 on average). The group with the lowest diA was mainly composed of CNM rams with no daughters in FRP (Table 1). As expected, both \( r_{GBV} \) and \( r_{PBV} \) decreased with the level of diA. There was no clear difference between \( r_{GBV} \) and \( r_{PBV} \) when diA was high. When diA declined the difference between \( r_{GBV} \) and \( r_{PBV} \) increased, as already observed by Clark et al. (2013) in meat sheep. Correlations of predictions of GBV and PBV with official EBV decreased as diA decreased. However, correlations between predictions of GBV and EBV were always higher than correlations between predictions of PBV and EBV (Table 1). From a practical point of view, overall results showed that the \( r_{GBV} \) of rams involved in the AI program, even with no daughters in FRP, are comparable to the accuracy of the parent average of young rams that currently enter the progeny test in the HB (0.52, unpublished data). For the CNM rams, the level of accuracy achieved by GBLUP, although lower than AI, is not far below that threshold (0.52) and may be improved by increasing the information on relatives in FRP.

Conclusions

Our results showed that genomic predictions of rams based on a female reference population can be used for selective breeding in dairy sheep. For traits difficult or costly to record, that are not routinely measured in the herd book, this approach may dramatically reduce the costs of performance recordings. This may largely counterbalance the additional genotyping costs. However, we must keep in mind that pedigree recording is crucial to makeup the female reference population in order to achieve an adequate level of information on relatives.

Acknowledgments

This study was funded by the project MIGLIOVIGENSAR, LR n.7/2007, R.A.S.

List of References


Table 1. Composition of the three relationship groups of rams, number of daughters in the female reference population, average accuracy of pedigree ($r_{PBV}$) and genomic ($r_{GBV}$) based predictions and correlation coefficients of pedigree ($PBV$) and genomic ($GBV$) based predictions with official $EBV$.

<table>
<thead>
<tr>
<th>Group</th>
<th>Ram s n</th>
<th>FRP sires$^1$ n</th>
<th>FRP daug.$^2$ n</th>
<th>$r_{PBV}$</th>
<th>$r_{GBV}$</th>
<th>cor(PBV,EBV)$^3$</th>
<th>cor(GBV,EBV)$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High relatedness</td>
<td>134</td>
<td>130</td>
<td>19.0</td>
<td>0.80</td>
<td>0.82</td>
<td>0.37</td>
<td>0.47</td>
</tr>
<tr>
<td>Medium relatedness</td>
<td>103</td>
<td>14</td>
<td>0.3</td>
<td>0.49</td>
<td>0.61</td>
<td>0.24</td>
<td>0.42</td>
</tr>
<tr>
<td>Low relatedness</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0.27</td>
<td>0.47</td>
<td>0.11</td>
<td>0.18</td>
</tr>
<tr>
<td>Whole sample</td>
<td>537</td>
<td>144</td>
<td>4.8</td>
<td>0.45</td>
<td>0.58</td>
<td>0.14</td>
<td>0.24</td>
</tr>
</tbody>
</table>

$^1$ number of rams used as sire in the female reference population
$^2$ average number of daughters per ram in the female reference population
$^3$ coefficient of correlation
Figure 1. Accuracy of genomic predictions ($r_{GBV}$) plotted against the measure of the information on relatives in the reference sample (diA).