

Predicted accuracies of GEBV from female nucleus and male reference populations in Sarda dairy sheep

Usai M.G., Salaris S., Casu S., Sechi T., Miari S., Carta P.,
Antonello Carta

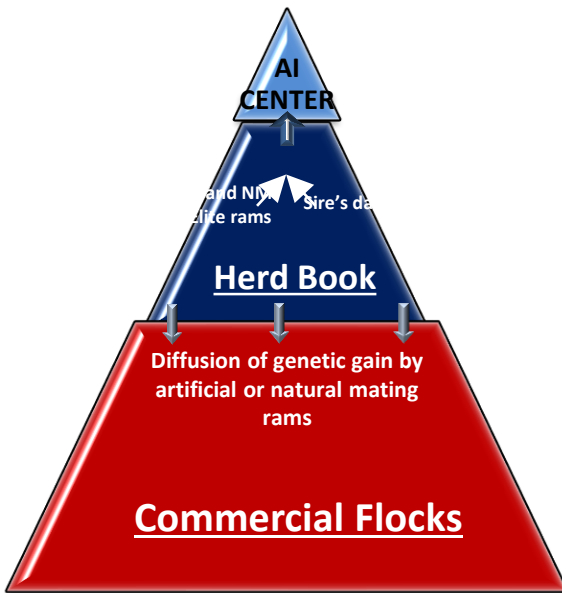


Agris Sardegna, Sassari, Italy

Acknowledgments

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

Balance of the selection scheme of the Sardinian breed



- **Low improvement of production traits**
 - Insufficient AI and single sire natural mating rates
 - Effective size of the selected population lower than the official
- **Few selected traits: low accuracy of performance records for animal health and functional traits**
 - lack of extensive historical recording, low heritability.

INSUFFICIENT INCREASE OF THE GLOBAL VALUE OF SARDINIAN SHEEP: risk of substitution of the local breed with exotic improved breeds (Assaf and Lacaune)

Genomic Selection

The basic idea of the genomic selection is to save time and money of performance recording respect to the traditional approach by:

- estimating effects of genome segments by dense marker arrays on just the genotyped and performance recorded portion of the selected population (reference population)
- producing GEBV for the genotyped selection candidates

Genomic Selection (dairy cattle approach)

Actually, the most successful application has been on dairy cattle by shortening the generation interval on the male selection pathways but without changing the traditional paradigms:

- large scale recording of traits for bulls reference populations
- wide use of elite bulls
- limited selection on the female pathways

Dairy sheep: technical issues

- **less accurate progeny test** due to the low average number of daughters , the lack of genetic connectedness and the low accuracy of the performance recording
- difficulty in finding **well-structured reference populations** to estimate SNP effects excepts some French breeds
- **accuracy is more relevant than generation interval**

Genomic selection in dairy sheep

is it feasible the dairy
cattle approach?

Dairy sheep: economic issues

- **high cost** of SNP arrays if related to the value of a single ewe and the number of recorded traits :
 - Cattle : 800.000 SNP x 10-27 traits
 - Sheep : 800.000 SNP x 1-5 traits (the effective number of estimated SNP effects is even lower when weighing for the accuracy of the performance records)

CAN WE DEVELOP A MODEL OF GENOMIC SELECTION SPECIFIC OF DAIRY SHEEP TAKING INTO ACCOUNT TECHNICAL AND ECONOMIC CONSTRAINTS?

The Sardinian approach



- organizing a female informative population in a “**genomic flock**” using AI and high impact HB rams
- performance recording for several health and functional traits
- realizing an optimal combining of medium density arrays genotyping with whole genome sequencing of the most influential animals on the genomic flock
- estimating accurately genome segment effects in the female nucleus and genotyping selection candidates rams of the HB with medium density arrays to produce GEBV
- detecting causal mutations and/or LD markers in the female nucleus

Aim of the study

Comparing the ability of the female informative population to provide accurate GEBV of the herd book selection candidates rams respect to a traditional male reference population

- Identifying a parameter estimating the **amount of the information on relatives** of selection candidates in the **2 reference populations**
- Using this parameter to predict the **expected accuracies** of the selection candidates **before genotyping**

Female informative population

experimental flock



3,949 ewes
were generated
by 161 rams.

Description:

- typical Sardinian dairy sheep farming system
- 1,000 milking ewes with a replacement of about 25% (from Sarda HB rams)

Management

- 40 daughters per ram until 2009
- 9 daughters per ram since 2010 (to increase the number of represented *bloodlines* from HB)

Measured traits

- Milk yield and composition, nutritional value
- milkability and udder morphology
- health (faecal egg count, ELISA test for paratuberculosis, ELISA test for visna-maedi)
- reproduction

Genotyping of all ewes and their sires (Illumina Inc. OvineSNP50 Beadchip, whole genome resequencing of more influential animals and imputation)

Available data

Female informative population (FN)

- 3,731 genotyped ewes
- 13,059 milk yield lactation records

Herd Book (sires) informative population

DYD calculated from the part of the recorded population which is actually exploitable for selective breeding (1,998,313 HB lactation records form 1983 to 2016)

- 39,067 daughters
(94 on average from 2 to 911)
- 100,114 milk yield lactation records
(242 on average from 2 to 2,494)

144 with daughters in FN

389 without daughters in FN

414 rams with daughters in the HB

119 rams without daughters in HB

533 genotyped rams

GEBV and rGBV calculation

Female informative population

$$y = Xb + Zg + Wp + e$$

y: lactation records

b: fixed effects

p: random permanent environmental effects

e: random residuals

X, Z & W: incidence matrices

G: included 3712 FN ewes and 533 rams

Herd Book (sires) informative population

$$y = \mu + Zg + e$$

y: DYD;

μ: mean;

e: random residuals;

Z: incidence matrix

G: included 533 rams

$$g = \mathbf{GEBV} \sim N(0, \mathbf{G}\sigma_g^2),$$

$\sigma_g^2 \rightarrow$ genetic variance

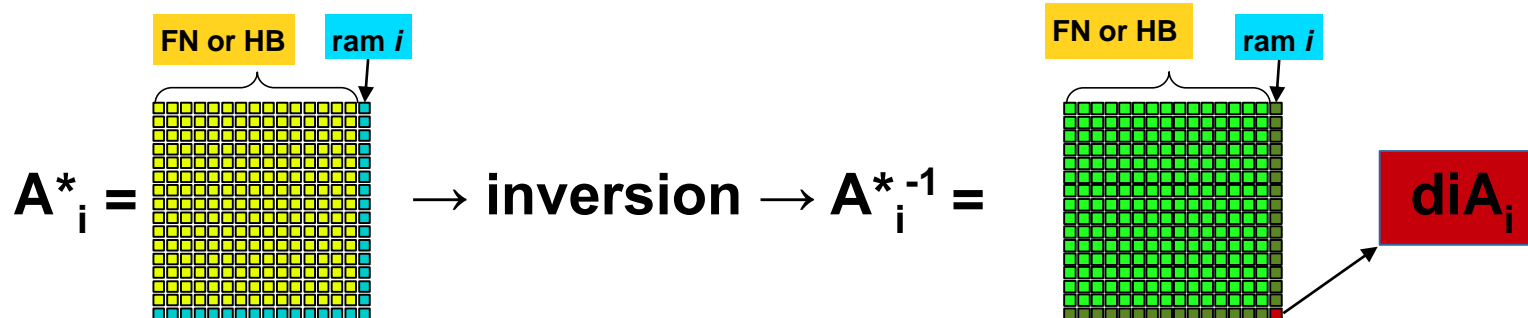
$$\text{theoretical accuracy of GEBV (rGBV}_i) = \sqrt{1 - \left(\frac{SE_i^2}{\mathbf{G}_{ii} \times \sigma_g^2} \right)}$$

Identifying a parameter estimating the amount of the information on relatives of selection candidates in the 2 reference populations

Numerator relationship matrix : $A_{FN} = 11,882$ individuals

$A_{HB} = 12,877$ individuals

diA_i : diagonal element corresponding to the target ram of the inversion ($A^*_{i^{-1}}$) of a section of the relationship matrix (A^*_i) including the ram itself and all individuals of the informative populations (**3731 FN** ewes or **414 HB** rams with DYD)



Correlating the identified parameter with the calculated GEBV theoretical accuracies to identify a function for predicting the expected accuracies of the selection candidates before genotyping

Female informative population

diA_i of 533 rams were plotted against the calculated theoretical accuracies ($rGBV_i$) and a predicting function was fitted (f_{FN})

Herd Book (sires) informative population

diA_i of 119 selection candidate rams were plotted against the calculated theoretical accuracies ($rGBV_i$) and a predicting function was fitted (f_{HB})

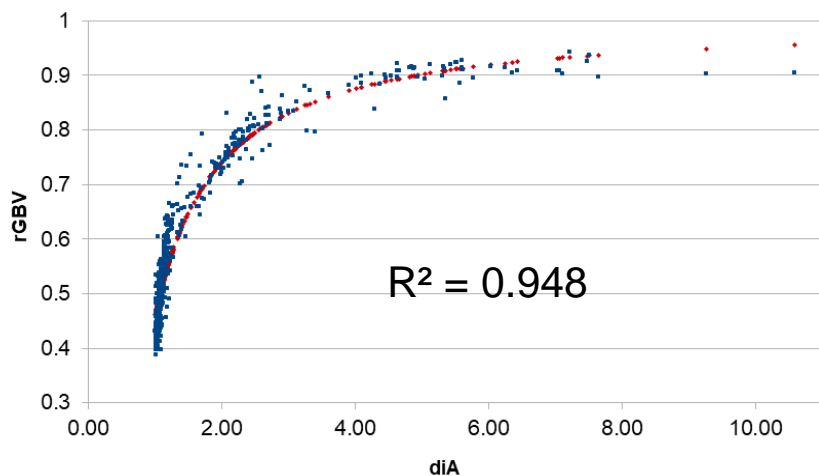
Predicting the expected accuracies for 310 HB selection candidates rams (year of birth 2013) and comparing the 2 populations

- f_{FN} on the diAi of the 310 rams from $A^*_i^{-1}$ with the 3371 ewes of FN
- f_{HB} on the diAi of the 310 rams from $A^*_i^{-1}$ with 4189 HB rams of a potential reference population i.e. rams born from 2003 to 2010 to include as many as possible of the sires of the selection candidates
- **Averages and distributions** of diAi and $E(rGBVi)$ were reported for the 2 populations

Correlating the information on relatives (diA_i) with the theoretical accuracies ($rGBV_i$)

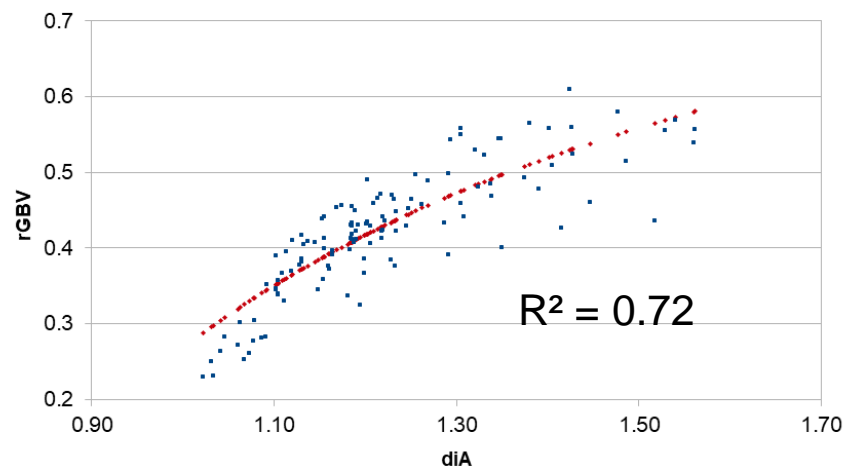
FN population:

Plot of diA_i against $rGBV_i$ (533 rams)



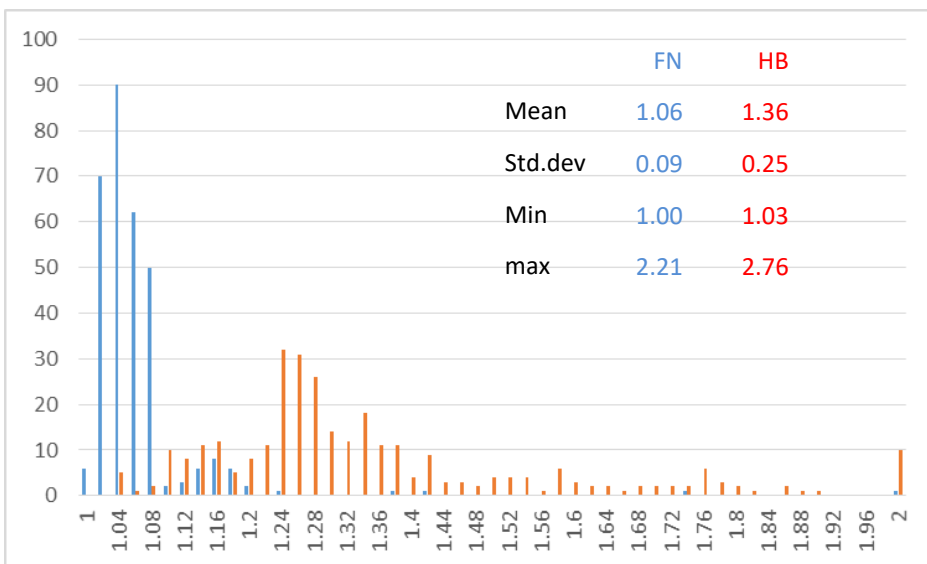
HB population:

Plot of diA_i against $rGBV_i$ and (119 rams without DYD)

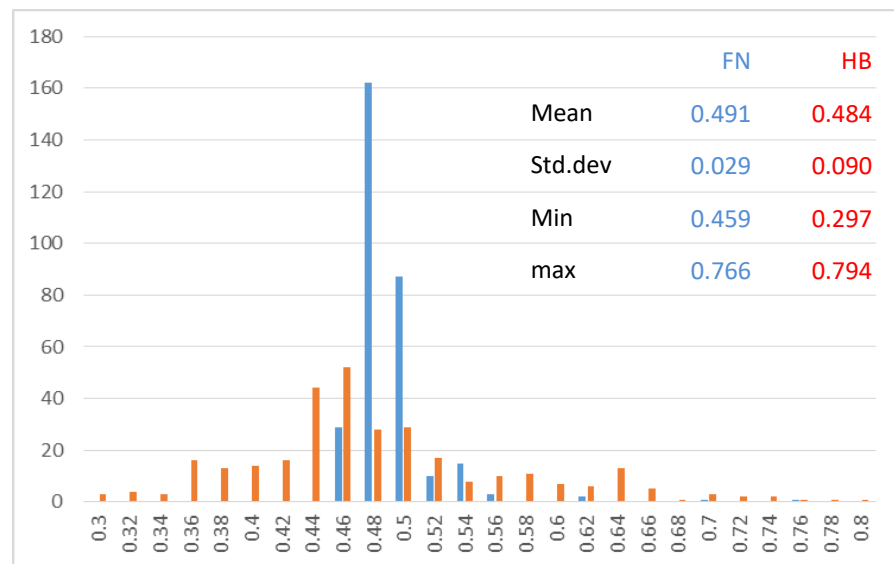


Predicting the expected accuracy $E(rGBVi)$ of 310 selection candidate HB rams

Distribution of diA_i derived from FN (blue) and HB (red)



Distribution of $E(rGBVi)$ derived from FN (blue) and HB (red)



Remarks

- The **numerator relationship matrix can be used to predict accuracies of genomic values of selection candidates** either from a female informative populations or a traditional male ancestors informative population
- Either the parameter measuring the information on relatives or the predicting functions of the **$E(rGBVi)$ can be improved**
- **The markedly different distributions of diA_i and $E(rGBVi)$ between the female and the male ancestors informative populations depend on the different kind of informative relatives (male ancestors vs female ancestors, daughters and half-sisters).**
- The **female informative population produces $E(rGBVi)$ of selection candidates as large as the male informative population** despite the lower amount of information on relatives

Remarks

In the Sarda sheep breed:

- **10 years of sires genotyping and daughters performance recording in HB would produce on average the same expected accuracies $E(rGBVi)$ than the female informative population but with a much higher variability**
- **Considering a replacement of 250 ewes per year in FN, the use of 50 rams with 5 daughters per ram will lead to an $E(rGBVi)$ of 0.58 for themselves and 0.54 for their sons (Salaris et al., 2018)**

PERSPECTIVES

In most dairy sheep breeds:

- **for many traits costly to measure on large scale, the female informative population is a realistic option** to produce (genomic) breeding values for HB selection candidate rams.
- **for traits routinely measured in HB such as milk yield, the female informative population is crucial to trigger** the application of the genomic selection allowing to produce GEBV since the starting of the genotyping program.
- **The progressive pile up of males genotypes** jointly to the application of **SSGBLUP methodologies** combining female nucleus and herd book data will allow to get **high accuracies of GEBV** which could not be obtained **using just the female or the male informative populations.**