



Analysis of a representative sample of Sarda breed artificial insemination rams with the OvineSNP50K BeadChip

M.G. Usai¹, T. Sechi¹, S. Salaris¹, T. Cubeddu²,
T. Roggio², Sara Casu¹, A. Carta¹

¹Research Unit: Genetics and Biotechnology, DIRPA-AGRIS Sardegna, Loc. Bonassai, 07100 Sassari, Italy.

²PortoConteRicerche Srl, Loc. Tramariglio, 07041 Alghero, Italy





OvineSNP50 BeadChip

- developed by Illumina in collaboration with ISGC
- 54,241 evenly spaced SNPs
- on average one marker per 46 kb





OvineSNP50 BeadChip : potential use

- **Genetic diversity and breed relationships**
- **Linkage disequilibrium**
- **Fine- mapping of QTL for Marker (Gene)
Assisted Selection**
- **Genomic selection**





OvineSNP50 BeadChip : recent applications

- ISGC : 24 individuals X 64 breeds to study genetic distances, phylogenesis, and LD
- Genomic selection in French Dairy Sheep (Astruc, this Icar session)
- Cross-breeding levels (Solkner, Interbull meeting)





This work

Evaluating the suitability of the OVINE SNP50K beadchip to the Sarda breed by:

- identifying not usable SNP
- calculating the information content and LD extent
- performing a preliminary association study with milk yield



Agris



Agenzia regionale per
la ricerca in agricoltura

MATERIALS & METHODS





Population

111 AI Sarda rams with a high genetic impact on the selected population born between 1985 and 1999

	Average	Std Dev	Min	Max
Daughters (N)	165	120	49	938
Lactations (N)	454	332	107	2583
Flocks (N)	75	42	14	246
Gran-Daughters (N)	929	2838	17	21773





Genotyping

- **DNA was extracted from frozen semen using a modified Heyen et al. (1997) protocol**
- **SNP genotypes were produced by the iScan System of Illumina (Porto Conte Ricerche)**



Data analysis

- Exclusion of not suitable markers
- Minor allele frequency distribution
- Observed and expected heterozygosities
- Extent of Linkage Disequilibrium (r^2) at different marker distances

$$r^2 = \frac{(fr(A_1B_1) \times fr(A_2B_2) - fr(A_2B_1) \times fr(A_1B_2))^2}{fr(A_1) \times fr(A_2) \times fr(B_1) \times fr(B_2)}$$



Association analysis

Phenotypes : de-regressed EBVs for milk yield

Statistical analysis : LASSO-LARS with cross-validations was chosen in order to select a subset of marker effects (small size of the analysed population)



(Least Absolute Shrinkage and Selection Operator)

- **Constrained version of Ordinary Least Squares (OLS)**
- **Minimizes the residual sum of squares constraining the sum of absolute values of the regression coefficients (t)**

$$\min \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^m x_{ij} \beta_j \right)^2 \right\} \text{ subject to } \sum_{j=1}^m |\beta_j| \leq t \text{ for } t > 0$$

- **The constraint (t) allows some estimated SNP effects to be exactly zero**



LARS

(Least Angle Regression algorithm)

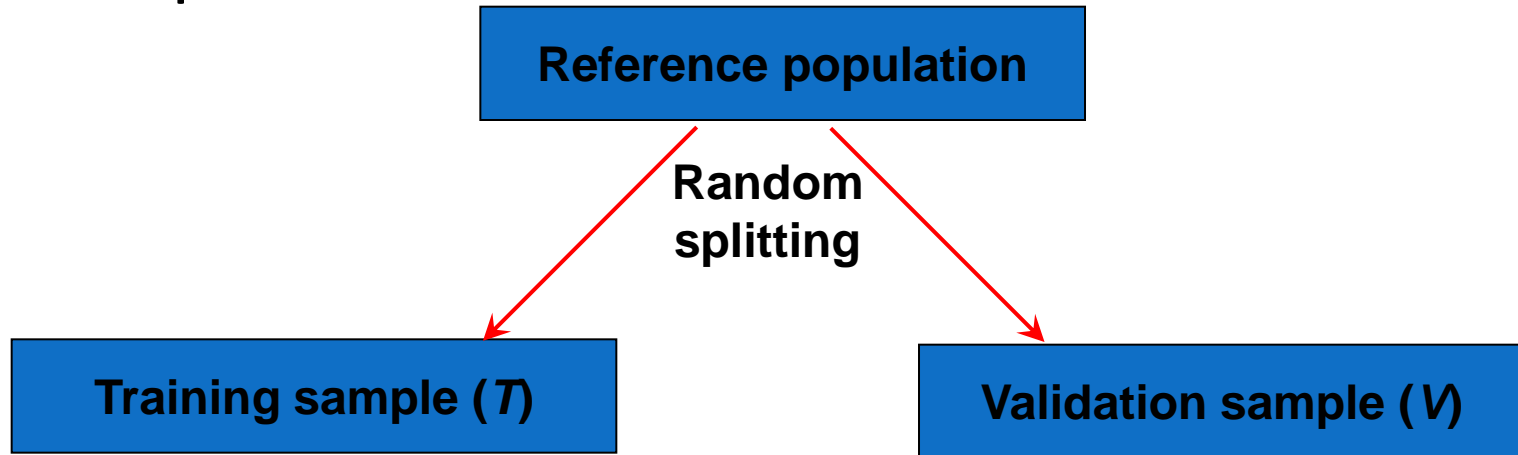
- **LARS estimates effects by successive iterations, for each iteration one new marker is added to the model**
- **The new marker is that with the highest absolute correlation between genotypes and current residuals**
- **Iterations stop when the maximum correlation between GEBV and phenotypes is reached**



Cross-Validations

- The best constraint ($\Sigma|\beta|$) estimation:

at each replicate:



Used to estimate the SNP
effects by LASSO-LARS

$T = 75\%$ of ref.pop

Used to validate the results obtained on T
maximum correlation $Y, GEBV$ - calculation of $\Sigma|\beta|$

$V = 25\%$ of ref.pop



Cross-Validations

- **1,000 replicates**
- **Best constraint (t)=mean of $\Sigma |\beta|$ over all replicates**
- **Selected SNP effects calculated on the whole population until the best constraint value is reached**
- **Frequency of occurrence over all replicates for each selected SNP**



Agris



Agenzia regionale per
la ricerca in agricoltura

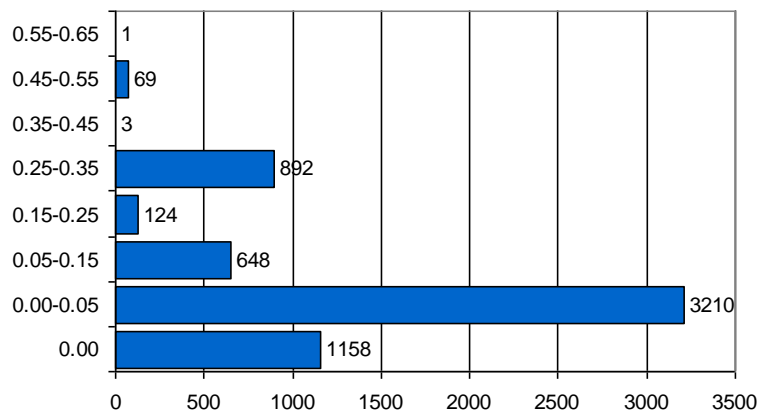
RESULTS



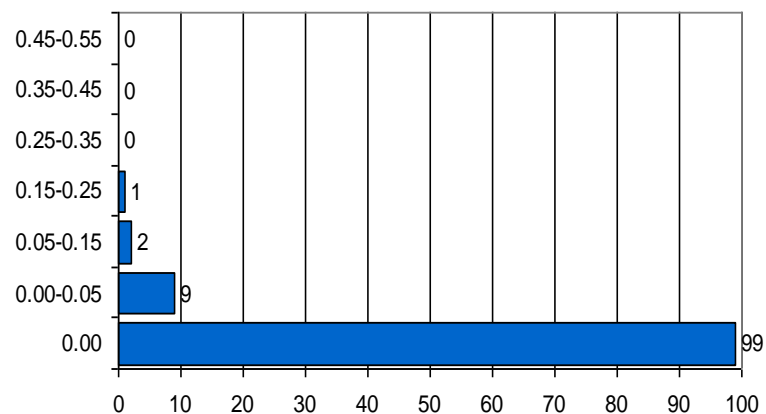


Population

	N	Av	SD	Min	Max
Inbreeding Coefficient (IC)	111	0.017	0.040	0.000	0.266
Relationship Coefficient (RC)	6105	0.065	0.102	0.000	0.578



Relationship Coefficient



Inbreeding Coefficient

All rams were successfully genotyped

Reasons for exclusion	N
N° no call > 0	2,783
unknown location	364
Chr X	1,341
HW disequilibrium ($\chi^2 < 0.05$)	4,950
MAF < 0.025	3,357

54,241 original SNP50K → 41,446 SNP retained



Map density

Length of the explored genome segment:

- **2,644,101 Kb \approx 2,644 cM**

Average distance between adjacent markers:

- **63.796 Kb \approx 0.064 cM**

Maximum distance between adjacent markers:

- **3,502.294 Kb \approx 3.5 cM**

Minimum distance between adjacent markers:

- **1.753 Kb \approx 0.002 cM**

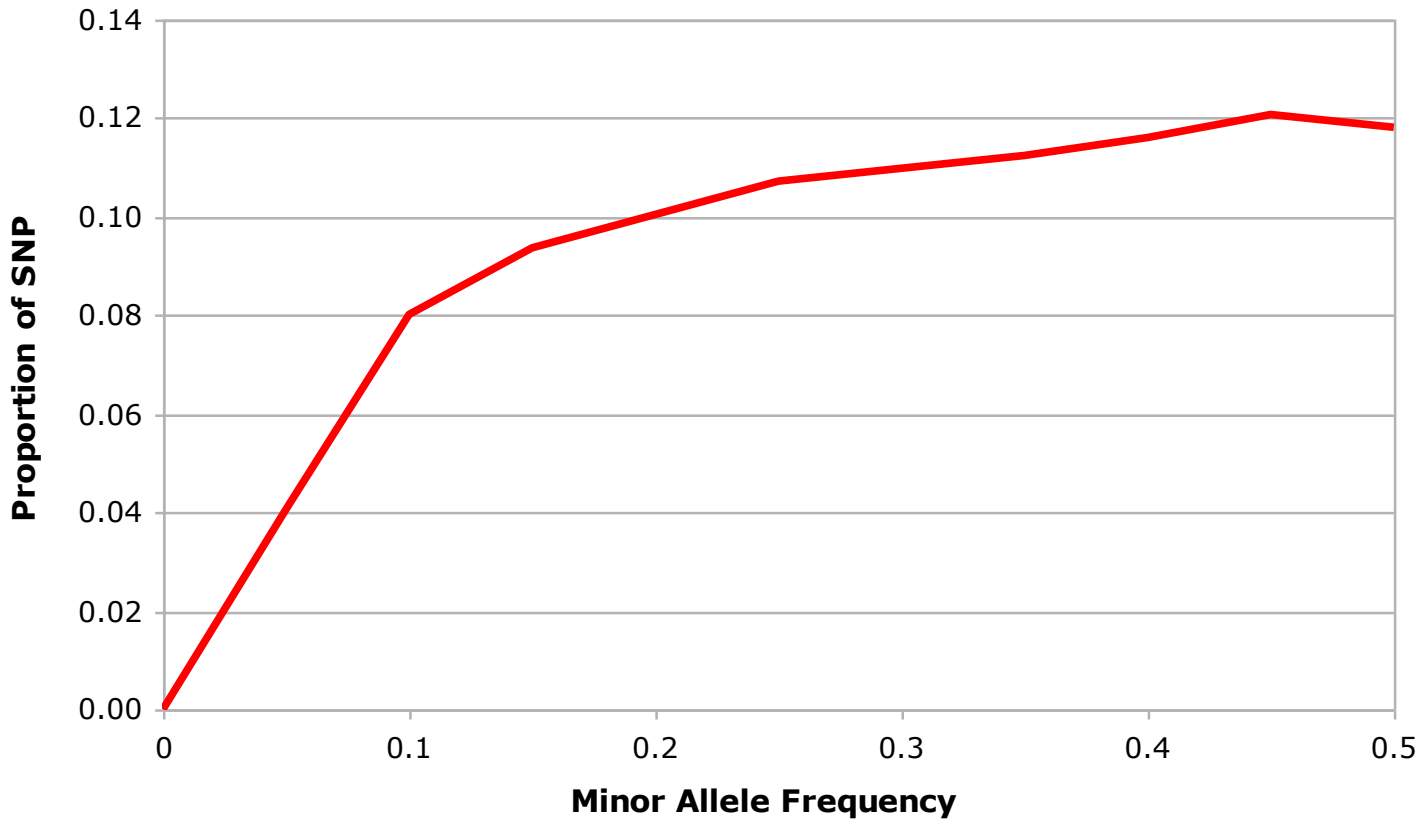


Agris Minor Allele Frequency (MAF)



Agenzia regionale per
la ricerca in agricoltura

average = 0.28





Heterozygosity

Average number of heterozygous SNP per ram :

- **More than one third (15,583)**

Observed heterozygosity:

- **0.38 ± 0.13**

Expected heterozygosity :

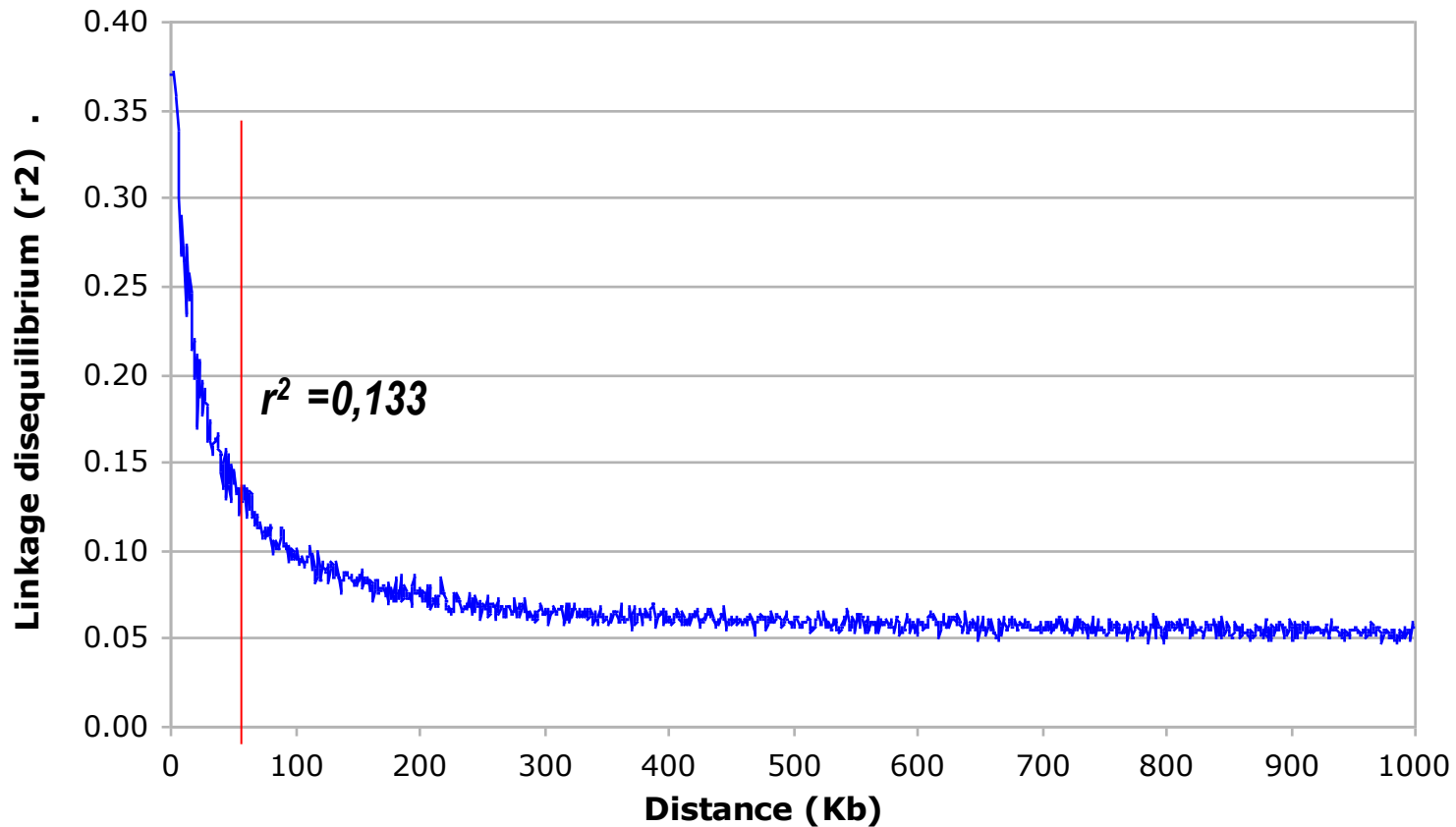
- **0.37 ± 0.13**





Linkage disequilibrium

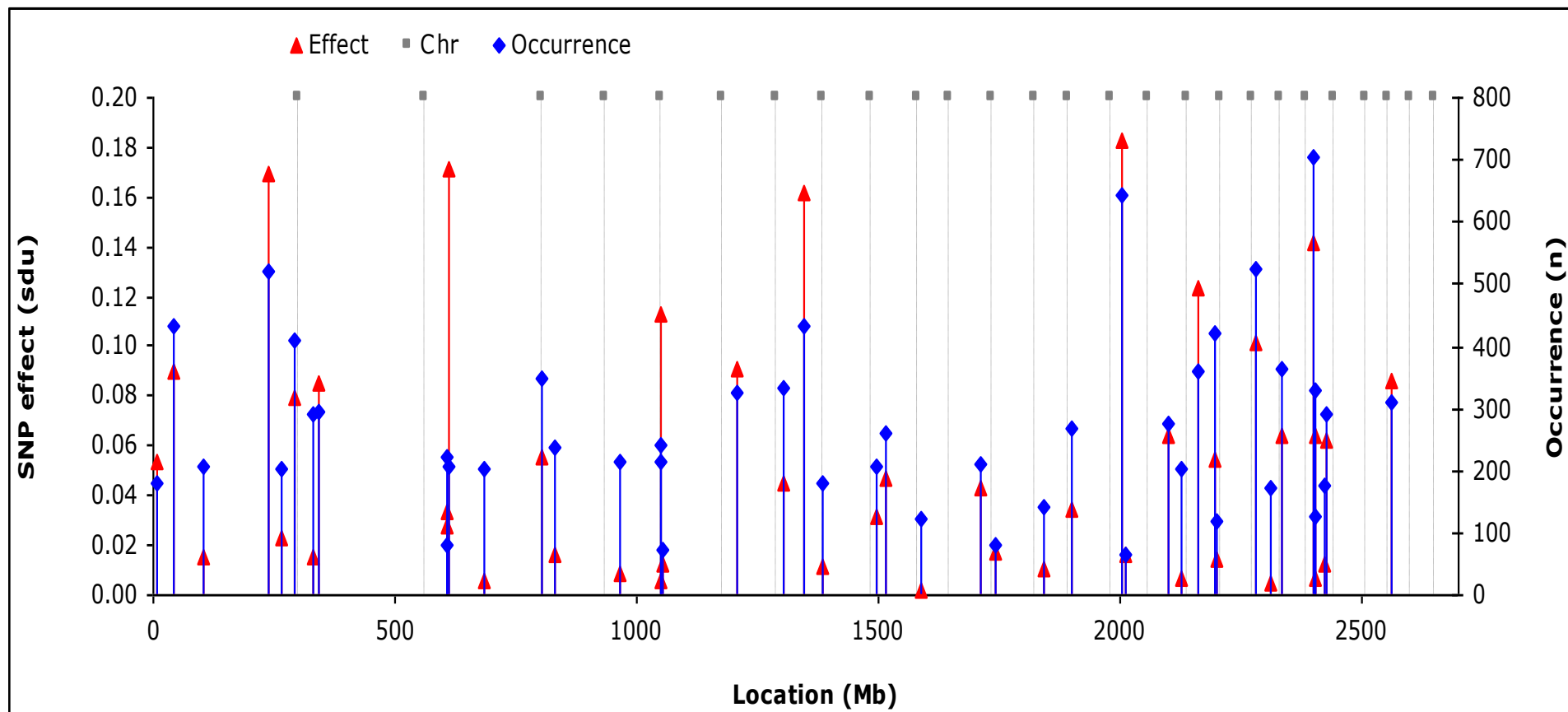
r^2 average value : 0.072





Selected SNPs effects

Best constraint (t) = 28.74 L; Selected SNP = 45; Explained variance = 0.72





CONCLUSION

Interesting perspectives for GWA studies

- Satisfactory genotyping results: 76% of total markers suitable for association studies
- High information content of the retained SNPs
- LD extent similar to other dairy sheep breeds (Churra and Lacaune) but lower than dairy cattle
- Small number of selected SNP spread all over the genome explaining a high portion of variance

