Analysis of a representative sample of Sarda breed artificial insemination rams with the OvineSNP50 BeadChip

M.G. Usai[1], T. Sechi[1], S. Salaris[1], T. Cubeddu[2], T. Roggio[2], Sara Casu[1] & A. Carta[1]

[1]Research Unit: Genetics and Biotechnology, DIRPA-AGRIS Sardegna, Loc. Bonassai, 07100 Sassari, Italy
[2]PortoConteRicerche Srl, Loc. Tramariglio, 07041 Alghero, Italy

Abstract

One hundred eleven artificial insemination rams with a high genetic impact on the Sarda breed selected population were chosen to be analysed with the OvineSNP50 BeadChip produced by Illumina. They had on average 165 lactating daughters in 75 flocks. The average relationship and inbreeding coefficients were 0.065 ± 0.102 and 0.017 ± 0.040 respectively. After editing 41,446 out of 54,241 available SNPs were retained. Reasons of exclusion were: number of no calls over 0 (2,783), unknown chromosome (364), X chromosome (1,341), significant deviation (P<0.05) from the Hardy-Weinberg equilibrium (4,950), minimum allelic frequency (MAF) lower than 2,5 % (3,357). The total length of the explored genome was 2,644,101 Kb (≈ 2,644 cM). The average distance between markers was 63.796 Kb (≈ 0.064 cM). The average MAF was 28%. The expected heterozygosity was on average 0.37±0.13. The average number of heterozygous SNPs per ram was 15,583. The LD on 1,000 Kb measured as the squared correlation (r^2) between pairs of loci was on average 0.07. Moreover, the LASSO-LARS procedure on the de-regressed EBVs for milk yield was performed. Forty five SNPs explaining 72% of the total variance were detected. The average SNP effect was 0.055 standard deviation units ranging from 0.002 to 0.183.

Keywords: SNP, Sarda dairy sheep, OvineSNP50 Beadchip.

1.0 Introduction

The recent availability of the OvineSNP50 Beadchip (Illumina) opens promising perspectives on using molecular information in the management of breeding schemes of dairy sheep populations. The large amount of information provided by this tool may have a strong impact on studies aimed at verifying the feasibility of Marker (or Gene) Assisted Selection or Genome-Wide Selection programs. The OvineSNP50 Beadchip has been recently used by the International Sheep Genomics Consortium (ISGC; *www.sheephapmap.org*) to genotype samples from 64 different sheep breeds. The Sarda dairy sheep breed population may have great advantages by introducing molecular information in the selection scheme. Indeed the possibility to predict breeding values by genomic data might lead to reduce costs of phenotype recording and increase the number of selection objectives. In the present work we analysed a sample of rams with a high genetic impact on the Sarda selected population with the OvineSNP50 Beadchip. Moreover, an association analysis aimed at detecting a set of SNPs affecting milk yield was performed.

2.0 Materials and methods

A sample of 111 Sarda rams born form 1985 to 1999 and with a high genetic impact on the selected population was chosen among those involved in the artificial insemination program of the Sarda breed. The lactating daughters per sire were on average 165 ranging from 49 to 938, distributed on average in 75 flocks. The rams were genotyped with the OvineSNP50 BeadChip produced by Illumina. Genotypes were edited to exclude SNPs not suitable for further analyses. The minimum allele frequency (MAF); the expected and observed heterozigosities and the extent of linkage disequilibrium (LD), measured by the squared correlation between pairs of loci (r^2; Hill and Robertson, 1968) were calculated. Moreover an association analysis between SNPs and de-regressed estimated breeding values (EBV) for milk yield was performed. In order to select a limited-size subset of SNPs the least absolute shrinkage and selection

operator (LASSO; Tibshirani, 1996) using the least angle regression (LARS; Efron *et al.*, 2004) algorithm, with cross-validations (Usai *et al.*,2009) was performed. At each cross-validation replicate the rams' population was randomly split into a training (T) and validation (V) sample. The T and V sizes were 75% and 25% respectively. The best constraint per cross–validation replicate was the sum of absolute effects corresponding to the LASSO-LARS step where the correlation between Genomic-EBV (calculated from the effects estimated in T) and phenotypes of V was maximized. One thousand cross-validation replicates were performed. Then the LASSO-LARS procedure was carried out on the whole population until the sum of absolute effects was equal to the mean of the best constraints over all cross-validations. Furthermore the frequency of occurrence was calculated as the number of replicates where a given SNP showed non zero effect.

3.0 Results

The average relationship coefficient between pairs of rams was 0.065 ± 0.102 and the average inbreeding coefficient was 0.017 ± 0.040. The average EBV was 20.9 ± 11.8. All rams were successfully genotyped. After editing 41,446 SNPs out of the original 54,241 were retained. In particular, reasons for SNP exclusion were: number of no calls higher than 0 (2,783); unknown location (364); chromosome X location (1,341), significant deviation (P<0.05) from the Hardy-Weinberg equilibrium (4,950); MAF lower than 2.5 % (3,357). The number of markers per chromosome ranged from 4,677 for OAR1 to 560 for OAR24.

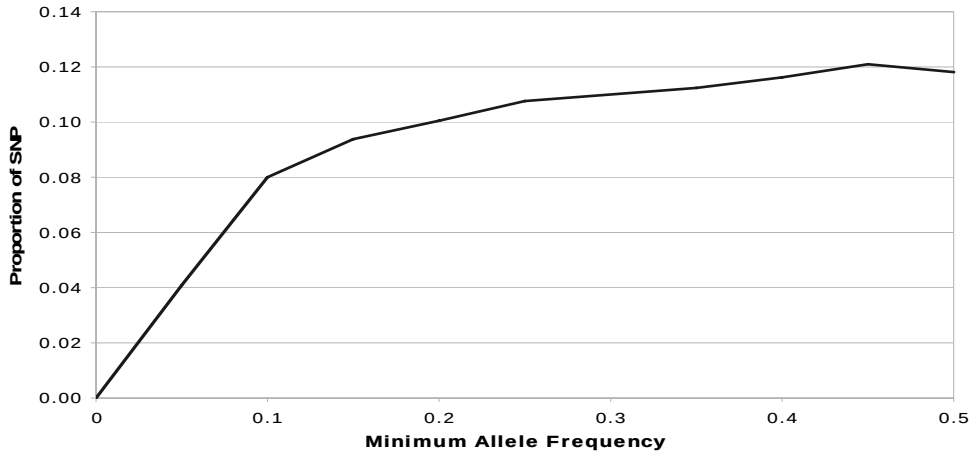

Figure 1. Minimum Allele Frequency (MAF) distribution.

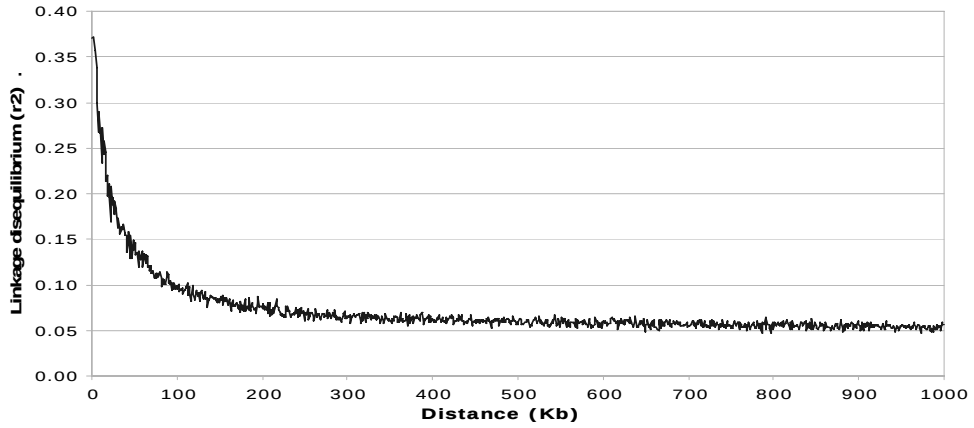

Figure 2. Linkage Disequilibrium (r^2) profile.

The total length of the explored genome segment was 2,644,101 Kb (≈ 2,644 cM). The average distance between flanking SNPs was 63.796 Kb (≈ 0.064 cM) with a maximum of 3,571 Kb (≈ 3.571 cM) on OAR10. The expected heterozygosity was 0.37 ± 0.13 on average and slightly lower than the observed one (0.38 ± 0.13). The average number of heterozygous SNPs per ram was 15,583. Figure 1 shows the

distribution of MAF. The mean and the mode (12%) were 0.28 and 0.45 respectively. Figure 2 shows the profile of the average r^2 according to the distance between SNPs. The maximum distance considered was 1000 kb (≈1 cM). The average value of r^2 over 1000 kb was 0.072. For a value of ≈64 Kb (average distance between retained SNPs) r^2 was moderate 0.133

As far as the association analysis is concerned, the overall best constraint estimated by 1000 cross-validations was 28.74 which corresponded according to LASSO-LARS procedure on the whole population, to 45 SNP with non zero effect. These SNPs explained 72% of the total phenotypic variance. Figure 3 shows the effects and the frequency of occurrence of detected SNPs and their location along the genome. The average SNP effect was 0.055 standard deviation units ranging from 0.002 to 0.183. The frequency of occurrence of the selected SNPs was on average 27% and ranged from 6.6% to 70%.

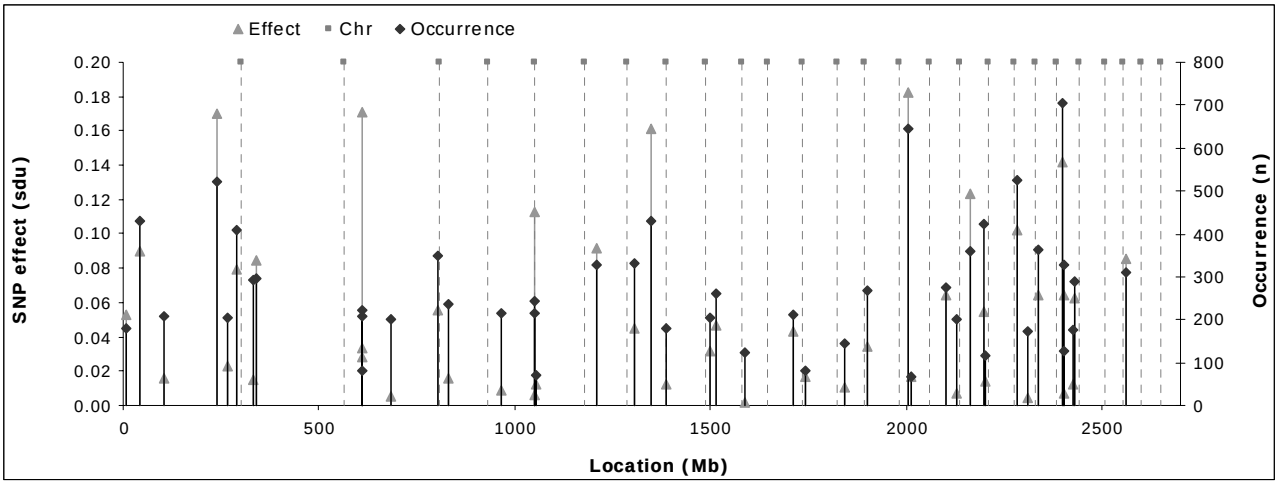

Figure 3. Effects and frequency of occurrence for the SNPs selected by LASSO-LARS procedure

4.0 Discussion and conclusion

Our results showed that the OvineSNP50 BeadChip can be successively used on the Sarda sheep breed. Indeed 76% of the original markers resulted suitable for genetic analyses. Furthermore, the MAF distribution and the observed heterozigosity showed a high information content of most markers. The r^2 profile was similar to those observed in other dairy sheep breeds as Churra and Lacaune (Raadsma *et al.*,2010). In Holstein dairy cattle for distances lower than 100 Kb, r^2 were higher than those estimated here (de Roos, 2008). This difference could be explained by lower selection pressures and larger effective population sizes in dairy sheep breeds. Concerning the association analysis results, even if they are based on a small sample of artificial insemination rams, the amount of explained variance and the frequencies of occurrence of the selected SNPs seem promising for future analysis based on larger samples.

This work was funded by the research program “APQ per la ricerca scientifica e l'innovazione tecnologica, progetto P5a – Attivazione del Centro di biodiversità animale per la valorizzazione del patrimonio animale con riferimento alla produzione e alla ricerca al servizio dell’allevamento” of the regional government of Sardinia.

5.0 References

de Roos, A.P.W., B.J. Hayes, R.J. Spelman, and M.E. Goddard, 2008. Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle. Genetics 179, 1503-1512.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004. Least angle regression. Ann. Statist. 32:407-499.

Hill, W. G., and A. Robertson, 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226-231.

Raadsma, H., 2010. LD analyses in the sheep HapMap project using an ovine 50K SNP array. http://www.sheephapmap.org/pag.php

Tibshirani, R., 1996 . Regression shrinkage and selection via the Lasso. J. Royal. Stat. Soc. B. 58:267-288.

Usai, M.G., M.E. Goddard, and B.J. Hayes, 2009. LASSO with cross-validation for genomic selection. Genet. Res. 91:427-436.