725. Estimation of linkage disequilibrium in Pag and Istrian sheep breed: towards genomic optimum contribution selection

A. Kasap[1*], J. Ramljak[1] and M. Špehar[2]

[1]University of Zagreb Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia; [2]Croatian Agency for Agriculture and Food, Svetošimunska 25, 10000 Zagreb, Croatia; akasap@agr.hr

Abstract

Knowledge of the genetic architecture of the population is important to predict the potential of genomic selection. The strength of association between genome wide single nucleotide polymorphisms (SNPs) and unknown quantitative trait loci (QTL), measured by linkage disequilibrium (LD), represents the corner stone of genomic selection. The aim of this study was to estimate LD in two Croatian native dual purpose (milk-meat) sheep breeds (Pag sheep, n=719 and Istrian sheep, n=749) using the Illumina Ovine 50K SNP chip. The main goal in these populations is to improve dairy yield traits, preferably via genomic optimum contribution selection. The estimated LD was low (r^2_{Pag}=0.04, D'_{Pag}=0.28, $r^2_{Istrian}$=0.06, $D'_{Istrian}$=0.39,) with very steep LD decay in both breeds. The results obtained are promising for conservation of these breeds, but less promising for the success of genomic selection. In order to maximize benefits of genomic information, single step genomic BLUP should be applied.

Introduction

Istrian and Pag sheep are Croatian native dairy orientated breeds with great economical, traditional and cultural importance. In order to preserve them '*in situ*', it is necessary to permanently increase their productivity. Small breeds under selection are under risk of genetic erosion, but low productivity leads them towards extinction due to non-competitiveness with highly selected breeds (FAO, 1998). Balancing between selection gain and loss of genetic variability is a challenging but feasible task, best achieved by following the principles of optimum contribution selection (OCS) (Wellmann *et al.*, 2021; Wooliams *et al.*, 2015). High-density SNP arrays have been used as an important source of information for selection (Van Raden, 2020) and conservation (Windig and Engelsma, 2010) in many livestock populations. However, benefits of genomic information in selection vary between populations, and depend heavily on population genetic parameters such as linkage disequilibrium (LD). Association between marker alleles (SNP) and causative alleles (QTL) directly affects the accuracy of genomic selection (Meuwissen *et al.*, 2001) and estimation of genetic diversity parameters (Boettcher *et al.*, 2010, Li *et al.*, 2011). Wishing to upgrade the existing selection strategy to genomic optimum contribution selection, we aimed to estimate the genome-wide levels of LD. The obtained results will serve as a base to ascertain the benefits of potentially novel selection strategy in these breeds.

Materials & methods

Samples for genotyping were collected with the permission of the Bioethical committee for protection and welfare of animals at University of Zagreb Faculty of Agriculture. Ear tissue from 719 Istrian sheep (13 flocks) and 749 Pag sheep (11 flocks) were taken with Allflex tissue sample unit applicator. Animals were originally genotyped for 52,152 SNPs, using the Illumina OvineSNP50K BeadChip. Only autosomal SNPs were included in the analysis after quality control parameters for missingness per animal, missingness per SNP, and minor allele frequency set to 0.9, 0.95, and 0.01, respectively. SNPs were also removed if they deviated from the Hardy-Weinberg equilibrium (*P*-value cut-off set to 1×10^{-6}). A total of 693 Istrian (44,743 SNPs) and 713 Pag sheep (46,685 SNPs) passed the above control. The linkage disequilibrium (LD) between pairs of SNPs was calculated with the PLINK (version 1.9) (Purcell *et al.*, 2007; Chang *et al.*, 2015) using the two most commonly used statistics r^2 and D'. Post hoc statistical analysis of genome wide LD was

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_725, © A. Kasap *et al.* 2022

conducted with dplyr (Wickham *et al.*, 2021) and ggplot (Wickham 2016) packages in R (R Core Team, 2020).

Results

The minimum and maximum chromosome-wide average LDs were: r^2_{min}=0.046, r^2_{max}=0.078, D'_{min}=0.314, D'_{max}=0.423 for Istrian; and r^2_{min}=0.027, r^2_{max}=0.049, D'_{min}=0.245, D'_{max}=0.307 for Pag sheep breed. The overall genome wide LDs were: $r^2_{Istrian}$=0.06, $D'_{Istrian}$=0.39, r^2_{Pag}=0.04, D'_{Pag}=0.28. The tabular (Table 1) and graphical representations (Figure 1) of LD change accompanied by between-marker distance revealed very similar pattern of LD decay between these populations. Graphical representation of LD decay obtained by plotting the average LD vs average between-marker distance of bins spanning 20 Kbp revealed a very steep decay of LD within first 150 Kbp, an intermediate decay from 150 to 250 Kbp, and negligible decay after 250 Kbp.

Table 1. Genome wide linkage disequilibrium (LD) classified by between-marker distance.

	Istrian sheep				Pag sheep			
Kbp	**mean (r^2)**	**sd (r^2)**	**mean (D')**	**sd (D')**	**mean (r^2)**	**sd (r^2)**	**mean (D')**	**sd (D')**
[0,100]	0.12	0.19	0.51	0.33	0.10	0.17	0.44	0.33
(100,200]	0.06	0.10	0.40	0.29	0.04	0.08	0.30	0.26
(200,300]	0.05	0.07	0.37	0.27	0.02	0.05	0.25	0.23
(300,400]	0.05	0.07	0.35	0.27	0.02	0.04	0.24	0.21
(400,500]	0.04	0.06	0.35	0.26	0.02	0.03	0.23	0.21
(500,600]	0.04	0.06	0.34	0.26	0.02	0.03	0.22	0.20
(600,700]	0.04	0.06	0.34	0.26	0.02	0.03	0.22	0.20
(700,800]	0.04	0.06	0.35	0.26	0.02	0.03	0.22	0.20
(800,900]	0.04	0.07	0.34	0.26	0.02	0.03	0.21	0.20
(900,1000]	0.04	0.05	0.33	0.26	0.02	0.03	0.21	0.20

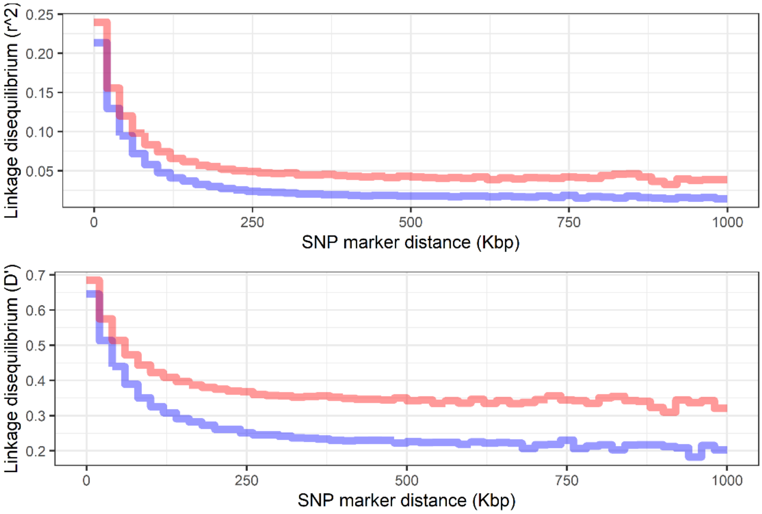

Figure 1. The linkage disequilibrium (LD) decay in Istrian (red) and Pag (blue) sheep.

Discussion

The extent of genome-wide LD and pattern of LD decay are important for GWAS studies, genomic selection, conservation genetics and many other applications. Therefore, our intention was to ascertain this parameter in Istrian and Pag sheep breeds and compare them with some other populations. In this study, we used the two most commonly used statistics, r^2 and D', to allow easy comparison of these results with some previous reports. However, in the between-study comparisons, we focused on r^2 due to its' greater robustness to allele frequency variation than D' (Ardlie *et al.*, 2002). Estimates of genome-wide LD and LD decay in both breeds were very low and steep, respectively. The estimates were very similar to those obtained for French Lacaune (Baloche *et al.*, 2014), Australian Merino, crosses Merino × Border Leicester, crosses Merino × Border Leicester × Poll Dorset (Al-Mamun *et al.*, 2015) and Chinese Merino (Liu *et al.*, 2017). The LD was considerably lower and LD decay less pronounced compared to pure Border Leicester and Poll Dorset breeds (Al-Mamun *et al.*, 2015). The results of the study by Alvarenga *et al.* (2018), conducted in Brazilian Santa Inês sheep, are also comparable to the estimates from this study, especially their reports for chromosome wide LD that were considerably lower (LD r^2_{min}=0.010, r^2_{max}=0.023, D'_{min}=0.176, D'_{max}=0.248). The low levels of LD and high rates of LD decay suggest high level of genetic variability in both breeds which is perspective from the conservation point of view, and less perspective for success of genomic selection. The required level of r^2 for genomic selection to achieve an accuracy of 0.85 should be 0.2 (Meuwissen *et al.*, 2001). The estimates of LD and LD decay have been reported in some other studies in sheep (e.g. Prieur *et al.*, 2017), but because of the different methodology and metric used in presentation of results it would be inappropriate to compare them with results from this research. The low LD in this study confirm many previous conclusions that sheep populations exhibit substantially higher genetic variability than cattle (e.g. De Roos *et al.*, 2008, Sargolzaei *et al.*, 2008). However, regardless of being less favorable than in dairy cattle, the estimated LD should be sufficient for accurate genomic selection (Goddard, 2009). To conclude LD profiles of the Istrian and Pag sheep indicate substantial genetic variability of the breeds and limited benefits of applying genomic selection in term of increased accuracy of estimated breeding values. Some objectively applicable strategies to bridge this gap in practice include genotyping with high density chips in combination with imputation techniques and implementation single step genomic approach in routine genetic evaluation system. Financial investments in novel selection strategies in these populations should be carefully examined to optimize selection under reasonable costs.

Funding

This research has been fully supported by the Croatian Science Foundation under the project Genomic characterization, preservation, and optimum contribution selection of Croatian dairy sheep (OPTI-SHEEP), grant number IP-2019-04-3559.

References

Al-Mamun H.A., Clark S.A., Kwan P., and Gondro C. (2015) Genet. Sel. Evol. 47(90). https://doi.org/10.1186/s12711-015-0169-6.

Alvarenga A.B., Rovadoscki G.A., Petrini J., Coutinho L.L., Morota G. *et al.* (2018) Sci. Rep. 8(8851). https://doi.org/10.1038/s41598-018-27259-7.

Ardlie K.G., Kruglyak L., and Seielstad M. (2002) Nat. Rev. Genet. 3:299-309. https://doi.org/10.1038/nrg777.

Baloche G., Legarra A., Sallé G., Larroque H., Astruc J.M. *et al.* (2014) J Dairy Sci, 97(2):1107-1116. https://doi.org/10.3168/jds.2013-7135.

Boettcher P.J., Tixier-Boichard M., Toro M.A., Simianer H., Eding H. *et al.* (2010) Anim. Genet. 41(1): 64-77. https://doi.org/10.1111/j.1365-2052.2010.02050.x.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. *et al.* (2015) Gigascience 4(1): s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8.

De Roos A.P.W., Hayes B.J., Spelman R.J., and Goddard M.E. (2008) Genetics 179:1503-1512. https://doi.org/10.1534/genetics.107.084301.

Food and Agriculture Organization of the United Nations (FAO) (1998) Secondary guidelines for development of National FAGR. Management of small populations at risk. Available at: https://www.portalces.org/sites/default/files/migrated/docs/smlpopn.pdf.

Goddard, M. (2009) Genetica 136(2):245-257. https://doi.org/10.1007/s10709-008-9308-0.

Li M.H., Strandén I., Tiirikka T., Sevón-Aimonen M.L., and Kantanen J. (2011) PLoS ONE 6(11): e26256. https://10.1371/journal.pone.0026256.

Liu S., He S., Chen l., Li W., Di J. *et al.* (2017) Genes Genomics 39(7):733-745. https://doi.org/10.1007/s13258-017-0539-2.

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819.

Prieur V., Clarke S.M., Brito L.F., and McEwan J.C. (2017) BMC Genetics 18(68). https://doi.org/10.1186/s12863-017-0534-2.

Purcell S.N.B., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R. *et al.* (2007) Am J Hum Genet 81(3):559-575. https://doi.org/10.1086/519795.

R Core Team (2020) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. Available at: https://www.R-project.org.

Sargolzaei M., Schenkel F.S., Jansen G.B., and Schaeffer L.R. (2008) J. Dairy Sci. 91:2106-2117. https://doi.org/10.3168/jds.2007-0553.

Van Raden P.M. (2020) J. Dairy Sci. 103:5291-5301. https://doi.org/10.3168/jds.2019-17684.

Wellmann R. (2021) optiSel: Optimum Contribution Selection and Population Genetics. R package version 2.0.5. 2021. Available online: https://CRAN.R-project.org/package=optiSel.

Wickham H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H., François R., Henry L., and Müller K. (2021) dplyr: A Grammar of Data Manipulation. R package version 1.0.7. Available online: https://CRAN.R-project.org/package=dplyr.

Windig J.J., and Engelsma K.A. (2010) Conserv. Genet. 11:635-641. https://doi.org/10.1007/s10592-009-0007-x.

Wooliams J.A., Berg P., Dagnachew B.S., and Meuwissen T.H.E. (2015) J. Anim. Breed. Genet. 132(2):89-99. https://doi.org/10.1111/jbg.12148.