Opportunities and obstacles of the use of genomic data in sheep breeding; large versus small populations

M. Špehar[1], J.M. Astruc[2], J. Ramljak[3] and A. Kasap[3]

[1]Croatian Agency for Agriculture and Food, Svetošimunska cesta 25, 1
0000 Zagreb, Croatia
[2]IDELE, Castanet-Tolosan, France
[3]University of Zagreb, Faculty of Agriculture, Department of Animal Science and
Technology, Svetošimunska 25, 10000 Zagreb, Croatia
Corresponding Author: marija.spehar@hapih.hr

Abstract

The objective of this paper is to draw the benefit and obstacles of using genomic data in sheep breeding through two contrasting situations: Croatia and France. Istrian (IS) and Pag (PS) sheep are Croatian native breeds undergoing selection on dairy traits (milk, fat, and protein) based on traditional pedigree BLUP. Genetic improvement of dairy performance via selection is important for their long term productive and economic viability, while maintaining the existing genetic variability is vital for their resilience in unpredictable future environment. Different SNP arrays have been used in many novel breeding programmes for the purpose of selection and conservation, and inclusion of genomic information in existing genetic evaluation of IS and PS seems to be promising and beneficial from both, the selection and conservation perspectives. The intention of the stakeholders involved in genetic evaluation of these breeds is to implement the basic principles of genomic optimum contribution selection (OCS) in existing breeding program in order to provide selection progress on targeted trait/s with minimal loss of genetic variability (ultimate goal of the OPTI-SHEEP project, CSF, IP: 2019-04-3559). So far, almost the whole breeding population of IS and PS (altogether ~4,000 animals) was genotyped using the Illumina OvineSNP50K BeadChip®. Transition from pedigree based BLUP to single-step GBLUP and development of pipelines for routine genomic OCS are still in progress. However, prior to major financial investments in genotyping and picking up the best strategy for long-term OCS, many practical and scientific questions need to be answered. Since the benefits of genomic selection heavily depend on population specific genetic parameters such as linkage disequilibrium, genetic connectedness between flocks, heritability of the traits, effective population size, etc., their estimation highlighted the potential benefits of genomic selection and some weaknesses that need to be worked on to maximize the benefits from this promising strategy. All these efforts will definitely impact long-term sustainability of Croatian sheep breeds which will benefit the breeders, consumers and corresponding (dairy) industries. By showing how to optimize selection with maintenance of genetic diversity, the results could also be beneficial to other sheep and livestock breeding programs. In France, all dairy sheep breeds have switched to genomic selection, with large reference population of rams for artificial insemination (Corse breed, n=0.4 K; Red – Faced Manech, n=3.5 K, Black – Faced Manech, n=0.7 K; Basco – Béarnaise, n= 1.2 K Lacaune breed, n= 7.4 K).

Each year, more than 8,000 new animals are genotyped with more and more affordable medium density SNP arrays (including all candidate rams and also some females in flocks with no AI nor control mating). In addition to single-step genomic evaluation, SNP markers are used for prediction of monogenic traits such as resistance to scrapie, and for parentage verification and discovery. The extra genetic gain achieved through the genomic program has already permitted to include novel traits, such as resistance to parasites, and will hopefully enable the inclusion of novel efficiency and resilience traits in the near future. As the French sheep populations are large, it is not feasible to genotype the whole female population, unlike the Croatian situation. In this respect, and through the comparison between those populations (large vs small populations, ancient vs recent breeding programs), we highlight similarities and differences in terms of opportunities and obstacles in the use of genomic data.

We also assess how an initiative to establish an EU Reference Centre (EURC) for performance testing and genetic evaluation in small ruminants could offer opportunities to better valorise genomic data, including across country evaluation, in more sheep breeds.

Keywords: genomic selection, genotyping, sheep, genetic variability,
Presented at: Session 9: Genomic's impact on Livestock Sustainability

Introduction

In traditional breeding methods animals are selected as parents for the next generations using breeding values estimated based on phenotypic information and genetic relationships (pedigree-based BLUP). In the last decade, genomic selection (GS) has been implemented in animal breeding by 'upgrading' traditional breeding methods with genomic information, i.e. genotypes of thousands of single-nucleotide polymorphisms (SNPs) covering the whole genome. Dairy breeding programs were the first to adopt GS showing many benefits (Schaeffer, 2006). In small dairy ruminants, implementation of GS is quite challenging compared to dairy cattle due to incomplete data, lack of large and excellent reference population, higher genetic diversity between and within sheep breeds, the cost of maintaining reference populations, constraints of fresh semen for artificial insemination (AI), and the higher costs of genotyping compared to the value of the animal (Van der Werf *et al.*, 2014). Despite these obstacles, genomic evaluation methods can increase the accuracy of estimation compared to traditional pedigree-based BLUP as reported in numerous dairy-oriented sheep breeding programs (Legarra *et al.*, 2014, Baloche *et al.*, 2014). A good strategy in designing reference populations in combination with efficient genotyping strategies gives possibility for well-designed breeding programs to apply cost effective GS. The benefits of GS heavily depend on population specific genetic parameters such as linkage disequilibrium (LD), genetic connectedness between flocks, effective population size, etc., and their estimation brought to light potential benefits of GS. Genomics can also be used as an important tool for assessment of genetic diversity of local sheep breeds providing accurate estimates of relationship between animals than pedigree records and avoiding inbreeding (Eynard *et al.*, 2015). This is particularly important in cases when accurate pedigree records are unavailable or incomplete which is often the case in sheep populations.

Sheep milk production is highly important in Mediterranean and Middle Eastern countries where most of the milk is processed into cheese of high quality. Consequently, milk recording is economically important to the dairy industry. According to the ICAR online survey, the number of European dairy ewes in official milk recording reached around 900,000 ewes in 2021 (Astruc *et al.*, 2023). More than half of these populations was recorded in France. On the other hand, countries with smaller dairy sheep populations like Croatia account less than 1% of milk recorded ewes. Both countries have used SNP arrays for the purpose of selection and conservation of dairy sheep breeds in their breeding programmes. The objective of this paper was to draw the benefits and obstacles of using genomic data in contrasting countries, i.e. large vs. small dairy sheep populations: France vs. Croatia. By comparing French (Lacaune, Red – Faced Manech, Black – Faced Manech, Basco – Béarnaise, Corsican) and Croatian (Istrian, Pag) dairy sheep breeds, similarities and differences in the feasibility of GS in dairy sheep breeds are highlighted.

Material and methods

Investments in breeding programs are often related to possibilities and improvements in:

- Collecting phenotypes.
- Prediction of breeding values.
- Dissemination of realised selection progress.

The main peculiarities of dairy sheep breeding programs in France and Croatia are disclosed in the next paragraph.

French dairy sheep population

Breeding programs are implemented for five dairy sheep breeds that have traditionally been farmed in three mountainous areas in southern France: The Lacaune breed (LAC) reared in Southern Massif Central (200,000 ewes in selection and 300 new AI rams each year), Red – Faced Manech (MTR), Black – Faced Manech (MTN), and Basco – Béarnaise (BB) reared in the Pyrenean mountains (altogether 120,000 ewes in selection and 250 new AI rams each year), and the Corsican breed (COR) reared in Corsica with (18,000 ewes in selection and 20 new AI rams each year). Breeding programs are designed on a pyramidal structure with the production flocks at the base and open nucleus flocks at the top. The latter is the basis for selection based on genetic and nowadays genomic evaluation having official milk and pedigree recording. The key point in breeding programs is usage of AI based on fresh semen (seasonal reproduction) mainly from the nucleus population (Larroque *et al.*, 2014; Astruc *et al.*, 2022). Before genomic era, conventional breeding schemes were based on progeny-tested AI rams. The AI rate in the flocks under selection ranges from 45% (COR) to 85% (LAC) and the number of yearly progeny-tested rams varied from 20 rams in COR breed to 300 in LAC breed (Astruc *et al.*, 2022). Genomic evaluation has been implemented in 2015 for the LAC breed followed by Pyrenean breeds (2017) and COR breed (2020). The reference populations (Table 1) are constituted of AI progeny-tested rams across the breed (from 0.4 K in COR to 7.4 K in LAC) with a depth getting back to the early 2000's. Each year, more than 8,000 new animals are being genotyped, including all candidate rams and some females in flocks without AI and control matings are performed with medium density SNP arrays.

The single step GBLUP method using the package BLUPf90 (Misztal *et al.* 2002) has been used for genomic evaluation. Genotypes are also used for predicting major

Table 1. Reference population sizes across breed in France.

Breed*	Genotyped rams	Genotyped AI rams with daughters	Rams genotyped yearly (2023)
LAC	36 K	7.4 K (since 1996)	~3,500
MTR	7.9 K	3.5 K (since 1998)	~700
MTN	1.4 K	0.7 K (since 1996)	~100
BB	2.8 K	1.2 K (since 1999)	~250
COR	2.8 K	0.4 K (since 2003)	~350

* LAC – Lacaune; MTR – Red – Faced Manech; MTN – Black – Faced Manech; BB – Basco – Béarnaise; COR – Corsican

genes (such as scrapie, SOCS2, horn) and for parentage verification and discovery of unknown parents. The extra genetic gain obtained with the genomic program has already permitted to include novel traits more easily, such as resistance to parasites, and will hopefully allow other novel traits (efficiency and resilience) to be included in the near future.

Creating an initial reference population is a large investment and therefore assistance from funding through research and development projects is required. In France four projects have been carried out to build the reference population and the engineering of the main breeds of small dairy ruminants.

Croatian dairy sheep population

Dairy sheep breeding in Croatia is concentrated in the Mediterranean part of the country (islands and hinterland). Two dairy breeds are included in the breeding programs: Istrian sheep (IST) reared in the Istrian peninsula (1,026 ewes and 35 rams in 12 flocks) and Pag sheep (PAG) reared in the island of Pag (4,086 ewes and 78 rams in 34 flocks). Milk recording started two decades ago, while BLUP genetic evaluation has been carried out for about a decade (CAA, 2017) based on test-day repeatability animal model (Špehar *et al.*, 2020). Unlike French breeding programs, there is no AI, which hampers establishment of genetic connectedness between flocks and consequently leads to bias in the ranking of animals' genetic merits (BVs) between the animals belonging to different flocks. The breeders regularly exchange rams between the flocks, but not systematically to exploit all the benefits of the across-flock BLUP. Furthermore, the population size is relatively small in both breeds, which leads to substantial level of inbreeding, so specially designed mating plans are required to minimize the inbreeding rate in both populations. Since the inbreeding is inevitable in the small population under selection, the Optimum Contribution Selection (OCS) approach appears to be the optimal selection strategy in this population to achieve genetic gain and mitigate the loss of genetic variability.

The future breeding strategies for these breeds are improvement of dairy performance through selection, which is necessarily for their long term productive and economic viability and maintenance of the existing genetic variability which is crucial for resilience in an unpredictable future environment. Implementation of the basic principles of genomic OCS in existing breeding program is a method of choice which enables selection progress on targeted trait/s with minimal loss of genetic variability.

So far, almost the whole breeding population of IST and PAG breeds (altogether ~4,000 animals) was genotyped (Table 2) with the Illumina OvineSNP50K BeadChip® for the purpose of selection and conservation. Since the Croatian breed populations are small, it is feasible to genotype the whole female population unlike the French situation where genotyping is dominantly based on males.

For the sake of higher accuracy, the existing genetic evaluation is planned to be upgraded to the single-step genomic BLUP (Legarra *et al.*, 2009). Development of pipelines for routine genomic evaluation based on BLUPF90, a family of programs is

Table 2. Number of genotyped animals by breed and sex in Croatia.

Breed*	Genotyped rams	Genotyped ewes	All
Istrian sheep	86	1,293	1,379
PAG sheep	140	2,543	2,683

in ongoing phase. Knowledge of the genetic architecture of the population is important to predict the potential of GS. Therefore, prior to transition to GS, especially due to additional financial investments, it is wise to examine some important population genetic parameters that reflect genetic diversity within and/or between these populations such as effective population size, inbreeding rate, genetic connectedness between flocks, linkage disequilibrium, etc. It is important not only to have some clues about the current state of the population under consideration, but also to set up the most promising strategy for OCS.

The main investment in the GS is the cost of genotyping, and so far several funding sources have been used for genotyping as follows: Scientific Project OPTI-SHEEP (CSF, IP-2019-04-3559) of Croatian Science Foundation, Rural Development Programme of the Republic of Croatia: Sub-Measure 10.2 – support for conservation and sustainable use and development of genetic resources in agriculture, and financial support by Breeding Association of Sheep and Goats.

Results and discussion

The French situation

In France, the annual genetic gain for economic index was estimated to be between 0.12 (BB) and 0.35 (LAC) genetic standard deviation. This observation is consistent when:

- Observing the evolution of the indices over several year.
- Calculating the gain realised each year to generate a new cohort through actual selection intensity, accuracy and generation interval.

As the selection indices are calculated to have the same variability of the milk yield, the genetic standard deviation is the one of milk yield in each breed. Based on the comparison between the genetic gain observed recently in genomic selection and the genetic gain observed several years ago before genomic selection, in the French sheep breeding schemes, Figure 1 showed that the extra annual genetic gains obtained in genomic selection vs traditional selection ranged from 16% (BB) to 57% (LAC). GBLUP resulted in greater accuracies of estimated breeding values (EBV) compared to pedigree-based BLUP although for some traits and population, the increase in accuracy was small. The gain in generation interval is limited. The key factor is the selection intensity realised for the choice of the males (which directly depends on the number of genotyped candidates).

In addition to single-step GBLUP, genotyping is used for prediction major genes and for parentage verification and discovery. Major genes were associated with various reproductive, disease or production traits of interest to breeders. The following major genes were identified and included in breeding programs: PRP (scrapie resistance), SOCS2 (susceptibility to inflammation of the mammary gland), and Horn (management of horn in MTR). As the costs of genotyping decreases, it is expected that new major genes like those for lethal mutations and cryptorchidism will be available in the near future. SNPs have proved to be useful to infer pedigree information and could be used both to detect misidentification of parents and to assign true parents among candidates. Parentage verification followed by parentage discovery has been applied on males chosen for genomic selection (before genomic pre-selection). For example, the proportion of false wrong sires in LAC and Pyrenean breeds was 4.6 and 4.4 and sire discovery was successful in 92% for LAC and in 87% for Pyrenean breeds. Parentage verification is of particular importance for sire discovery in flocks that does not use AI (including organic farms where synchronisation by hormones is forbidden). The extra genetic gain obtained with GS will be used for inclusion of new resilience (functional

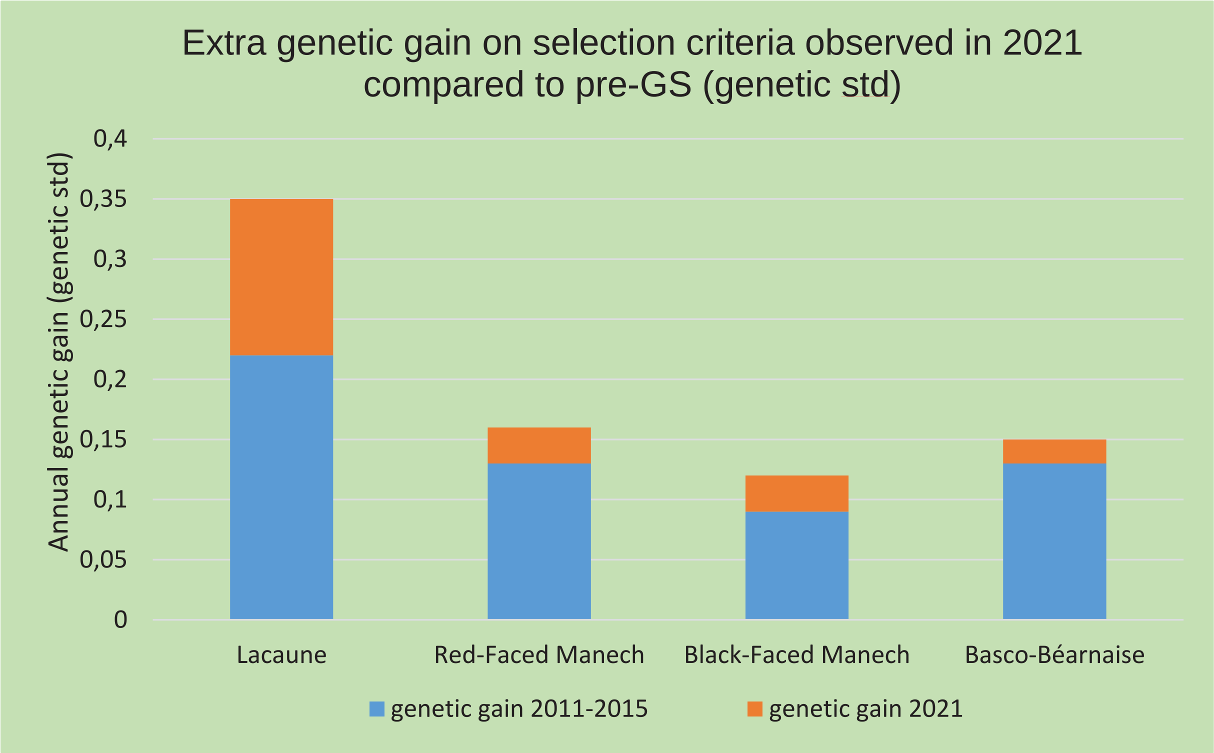

Figure 1. Genetic trends (in genetic standard deviation or σ_g) in the French dairy sheep breeds in GS (blue+orange) and comparison with pedigree-based BLUP (in blue).

longevity, resistance to parasites) and efficiency (proxies of feed efficiency) traits in the next years (Astruc *et al.*, 2022).

The Croatian situation

In Croatia, in order to provide basic information essential for designing the selection strategy in the IST breed, genetic connectedness between flocks have been examined (Kasap *et al.*, 2021). The success of across-flock genetic evaluation system depends on the genetic connectedness between flocks. When flocks are sufficiently connected, the BLUP genetic evaluation is robust, and EBVs can be fairly compared between flocks. On the other hand, limited connectedness leads to bias when comparing EBVs of animals belonging to different flocks (Kuehn *et al.*, 2008). Several statistical measures have been used to examine the degree of connectedness as follows: prediction error variance of differences in EBVs between animals belonging to different flocks, variance of estimated differences between management units, coefficient of determination of the difference between predicted breeding values, and correlation between predicted breeding values of individuals from different flocks. These statistics are useful to estimate the risk of comparing EBVs between flocks, as well as to design breeding schemes aimed at effectively linking flocks. Results showed that despite being useful in detecting flocks that tend to share breeding animals more frequently and vice versa, results (PEVD and CD) were insufficiently informative to tell us more about the bias in ranking of the animals EBVs. Connectedness based on *r*-connectedness statistic was below the acceptable level for unbiased ranking of animals belonging to different flocks (*ri,j* = 0.05). In order to increase the link between flocks in this population, some of the long-term specially designed breeding schemes need to be implemented. The best-known schemes to provide connections between different management units are the rotation of rams between herds (circle rams) and the sire references scheme.

In the Croatian case, population genetic parameters that reflect genetic diversity such as linkage disequilibrium (LD, effective population size, and inbreeding - F_{ROH}) have been examined with genomic data in IST and PAG breeds. LD was estimated because of its influence on the expected accuracy of genomic predictions (Baloche *et al.*, 2014). The estimated LD was low (r^2 Pag = 0.04, *D'*Pag = 0.28, r^2 Istrian = 0.06, *D'*Istrian = 0.39,) with a very steep LD decay (Figure 2) in both breeds (Kasap *et al.*, 2022). The results obtained are promising for conservation of these breeds, but less promising for the success of GS. To maximize benefits of genomic information, single step genomic BLUP should be applied for genomic evaluation.

Effective population size and inbreeding (F_{ROH}) have been also studied in both breeds. There was strong evidence of recent inbreeding, and the estimate of genomic inbreeding (F_{ROH}) in IST ($F_{ROH>2}$ = 0.062) was twice as high as in PAG ($F_{ROH>2}$ = 0.029) and was on the edge of acceptable levels (Ramljak *et al.*, 2024). Therefore, optimisation of mating plans is needed in the future to maintain genetic variability. Estimates of effective population size (N_e) evidenced more genetic variability in the PAG (N_e = 838) compared to the IST (N_e = 197) breed. Based on the average F_{ROH} of different size of ROH segments ($F_{ROH2–4}$, $F_{ROH4–8}$, $F_{ROH8–16}$, and $F_{ROH>16}$) a high contribution of recent inbreeding in the overall inbreeding is evident (Ramljak *et al.*, 2024). Signatures of selection were found only in IST and were associated with growth, feed intake, milk production, and immunity traits. The results of the studied population genetic parameters showed an unfavourable loss of genetic variability in the Croatian sheep population

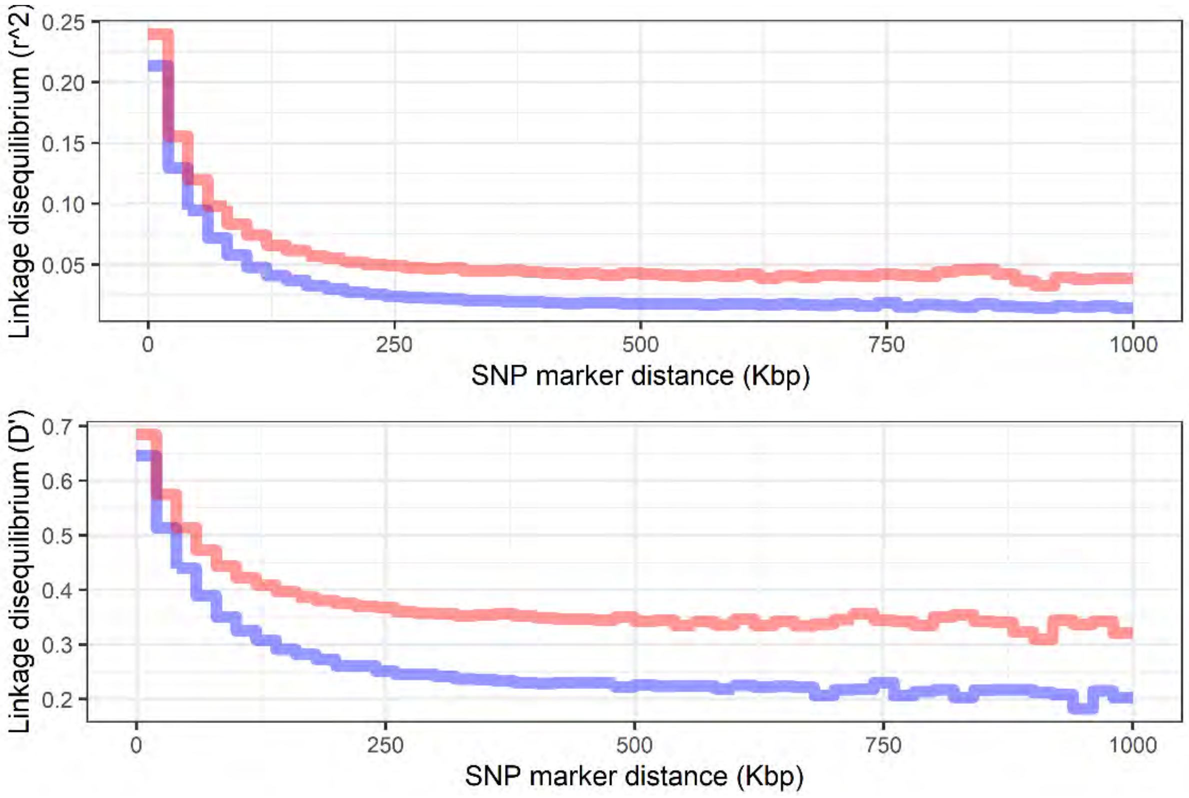

Figure 2. The linkage disequilibrium (LD) decay in IST (red) and PAG (blue) sheep

which is common in small population under selection. There is an urgent need for action to slow down this process and to create the necessary conditions (connectedness) to successfully utilize the entire genetic pool of the breed in a long-term selection.

An international initiative

It is important here to mention as well an initiative at the international level, driven by ICAR, to improve genomic evaluations in multi-breed context by establishing an EU Reference Centre (EURC) which would be responsible for performance testing and genetic/genomic evaluation in small ruminants. The EURC could offer opportunities for better valorisation of genomic data, including across country evaluation and evaluation of more sheep breeds. Across country evaluation in some multi-country breeds already exists such is the case in the French Manech and the Spanish Latxa (ARDI project). Beside evaluation, the EURC could also ensure a better exchange of genotypes for parentage verification and for share expertise across sheep and goat breeds.

Conclusion

Implementing genomic selection in sheep breeding in France has shown promising outcomes. Genomic selection not only accelerated genetic gain by allowing breeders to select animals with superior traits at an early age and based on more accurate breeding values, but also facilitated the identification of carriers of genetic markers associated with disease resistance (scrapie and mastitis) and the absence of horns. On the other hand, in Croatia, genomic selection has not yet been implemented in practice, and the population parameters that have been estimated indicate that there are some obstacles that must be overcome in order to fully exploit the potential benefits of this advanced genetic evaluation system. Following the successful example of France, it can be concluded that genomic selection has potential for improving productivity, sustainability, and profitability of Croatian sheep populations as well, but hard work is needed to strengthen connectedness between the flocks that will be included in the joint across-flock evaluation system in the future.

List of references

Astruc, J.M., Lagriffoul1, G., Legarra, A. and Buisson D. 2022. Six years of genomic selection have increased the genetic gain in French dairy sheep. Proceedings of 12th World Congress on Genetics Applied to Livestock Production. Rotterdam, Netherland, 2960-2963.

Astruc, J.M., Carta, A., Simčič, M., Špehar, M., Ugarte, E., Mosconi C. and Sievert, S. 2023. Milk recording in sheep and goats: state of the art and materials used for recording and sampling. Proceedings of the 46th ICAR Annual Conference held in Toledo, Spain, 21 – 26 May 2023. ICAR Technical Series no. 27

Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.M., Robert-Granié, C. and Barillet, F. 2014. Assessment of accuracy of geomic prediction for French Lacaune dairy sheep. J. Dairy Sci. 97:1107–1116.

CAA (Croatian Agricultural Agency). Annual Report for Year 2016. 2017. Croatian Agricultural Agency: Zagreb, Croatia, 2017.

Eynard, S.E., Windig, J.J., Leroy, G., van Binsbergen, R. and Calus .P.L. 2015. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. BMC Genet. 16:24.

Kasap, A., Ramljak, J. and Špehar, M. 2021. Estimation of Population-Specific Genetic Parameters Important for Long-TermOptimum Contribution Selection - Case Study on a Dairy Istrian Sheep Breed. Animals, 11, 2356.

Kasap, A., Ramljak, J. and Špehar, M. 2022. Estimation of linkage disequlibrium in Pag and Istrian sheep breed: towards genomic optimum contribution selection. Proceedings of 12th World Congress on Genetics Applied to Livestock Production. Rotterdam, Netherland, 2988 - 2991.

Kuehn, L., Notter, D., Nieuwhof, G.J. and Lewis, R. 2008. Changes in connectedness over time in alternative sheep sire referencing schemes. J. Anim Sci. 86: 536–544.

Legarra, A., Aguilar, I. and Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92: 4656–4663.

Larroque, H., Barillet, F., Baloche, G. and Astruc J.M. 2014. Toward genomic breeding programs in French dairy sheep and goats. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production (WCGALP) (Species Breeding: Sheep and goats), p. 1-6., Vancuver, Canada, 332.

Legarra, A., Baloche, G., Barillet, F., Astruc, J.M., Soulas, C., Aguerre, X., Arrese, F., Mintegi, L., Lasarte, M., Maeztu, F. and *et al.* 2014. Within and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. J. Dairy Sci. 97:3200–3212.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T. and Lee, D.H. 2002. BLUPF90 and related programs (BGF90). Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.

Ramljak, J., Špehar, M., Ceranac, D., Držaić, V., Pocrnić, I., Barać, D., Mioč, B., Širić, I., Barać, Z., Ivanković, A. and *et al.* 2024. Genomic Characterization of Local Croatian Sheep Breeds-Effective Population Size, Inbreeding and Signatures of Selection. Animals, 14, 1928.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim Breed and Genet. 123: 218–223.

Špehar, M., Mulc, D., Barać, Z., Mioč, B. and Kasap, A. 2020. Estimation of genetic parameters for dairy traits in Pag sheep with single and multi-trait test-day models. Small Rum Res. 183, 106029.

Van der Werf, J.H.J., Banks, R.G., Clark, S.A., Lee, S.J., Daetwyler, H.D., Hayes, B.J. and Swan, A.A. 2014. Genomic selection in sheep breeding programms. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production (WCGALP) (Species Breeding: Sheep and goats), p. 1-6., Vancuver, Canada 351.