

# Geographic Patterns of Genetic Variation in Indigenous Eastern Adriatic Sheep Breeds

---

Maja BANADINOVIĆ<sup>1</sup>

Alen DŽIDIĆ<sup>1</sup>

Mojca SIMČIČ<sup>2</sup>

Dragica ŠALAMON<sup>1</sup> 

---

## Summary

The geographical patterns of genetic variation in domestic species are the outcome of spatially explicit demographic events. Technological advances have made the obtaining of spatial information easier and different techniques have been developed to analyse spatial patterns of genetic variation. Relatively weak geographic trends found in sheep populations suggested the use of more powerful approaches, such as spatial principle component analysis. The aim of the study was to assess if application of spatial approach could reveal structures and patterns of genetic variation in indigenous Eastern Adriatic sheep breeds and contribute to our current knowledge about their genetic differentiation. We found south-east to north-west cline as the global structure using spatial principal component analysis, which outperformed the principal component analysis in this study and complemented understanding of variability of the investigated breeds reported in other multivariate and model-based clustering methods applied on this microsatellite data in previous research. Kernel density estimation suggested the Lika pramenka sheep belongs to a separate patch, not recognized in the spatial effects of the spatial principal component analysis. This potential structure should be further investigated.

---

## Key words

spatial genetics, sheep breeds, spatial principal component analysis, Moran's I

<sup>1</sup> Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetošimunska 25, 10000 Zagreb, Croatia

✉ e-mail: [dsalamon@agr.hr](mailto:dsalamon@agr.hr)

<sup>2</sup> University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia

Received: May 14, 2017 | Accepted: August 16, 2017

## Introduction

The geographical patterns of genetic variation in a domestic species are the outcome of demographic events of domestication, migration, selection, isolation, and expansion of successful breeds, which are by themselves also spatially explicit (Manni et al., 2004). As technological advances have made obtaining spatial information easier, there is a strong interest to include them in the analysis of genetic data (Joost et al, 2010; Joost et al., 2015). The geographic mosaic of genotypes that reflects the movement of alleles, the impact of genetic bottlenecks, and the effects of local selection pressures (spatially and temporally) could be used to infer evolutionary history and the processes that have shaped these geographic patterns in domestic animals (Laloë et al., 2010), for conservation purposes (Ruiz-Garcia and Jordana, 2000) or for learning about the adaptation of domestic animals to different environments (Pariset et al., 2012). In domestic animal conservation, breed is usually the object of conservation efforts, and spatial data is proposed for integration in the analyses of conservation of farm animal genetic resources (Duruz et al., 2017), even though the guidelines for genetic characterization of genetic resources only state its importance in visualisation (FAO, 2011.)

Different techniques have been developed to analyse spatial patterns of genetic variation among populations exploiting the geographic dimension of genetic data (François and Waits, 2016; Manel et al., 2003). Spatial auto-correlation analysis (Smouse and Peakall, 1999) is the initial approach to detect if proximate entities are more or less similar than expected for a random distribution (Ruiz-Garcia and Jordana, 2000). Mantel correlogram detects spatial structuring, without providing visualization of the spatial patterns (Legendre and Legendre, 1998). Different model based clustering approaches are inappropriate strategy when individuals are genetically structured as a cline (Jombart et al., 2008). Widely used exploratory alternatives are reduced space ordination methods, such as principle component analysis (PCA) (Patterson et al., 2006). Spatial multivariate methods, namely spatial principle component analysis (sPCA) and spatial multidimensional scaling (sMDS), are recommended tools for describing the correlation of genetic variation of domestic animals with geography by Laloë et al. (2010). Unlike PCA, they

use explicit geographic information, location coordinates. Global structures (patches, clines, intermediates) are distinguished from local ones (strong genetic differences between neighbours) and from random noise. Different patterns can be described in the revealed structuring: isolation by distance, clines, metapopulations and barriers to gene flow.

The aim of this study was to assess if application of spatial approach, namely sPCA, could reveal structures and patterns of genetic variation in indigenous Eastern Adriatic sheep breeds better than PCA, and contribute to our knowledge about their genetic differentiation.

## Material and methods

Obtaining the genotypes on 28 microsatellite loci (392 alleles) of 317 animals was described in Salamon et al. (2014), where names of the markers and the diversity parameters for the markers are reported. Genotype data were converted into allele counts per population/breed. Animals were sampled from 13 populations described in supplementary material of Salamon et al. (2014) and by Šalamon et al. (2015). Breed names, countries of origin, sample group abbreviations and geographical coordinates are given in Table 1.

Calculations and visualizations were obtained using the R software (R Core Team, 2011). First PCA, and then sPCA were performed with adegenet package (Jombart, 2008). Function scaleGen was used to scale and centre the data. Kernel density estimation on the three retained principal components was performed to visualize structure identified by PCA and sPCA. In Moran's I test, performed in spdep package (Bivand and Piras, 2015), and in sPCA, we used k-nearest neighbours connection network ( $k=4$ ). Spatial structures detected by sPCA were tested using the global and local permutation tests using Moran's eigenvector maps (Laloë et al., 2010). The confirmed spatial structure is visualized on the altitude map of the sampling area in order to assess spatial patterns. Visualizations were performed using adegenet, maptools (Bivand and Lewin-Koh, 2017) and raster (Hijmans, 2016) packages. We checked if the first two retained axes from PCA and sPCA were correlated. Also, spatial autocorrelation of the first three PCA axis was tested using Moran's I test.

Table 1. Thirteen sampled populations with sample abbreviations and locations of sampling

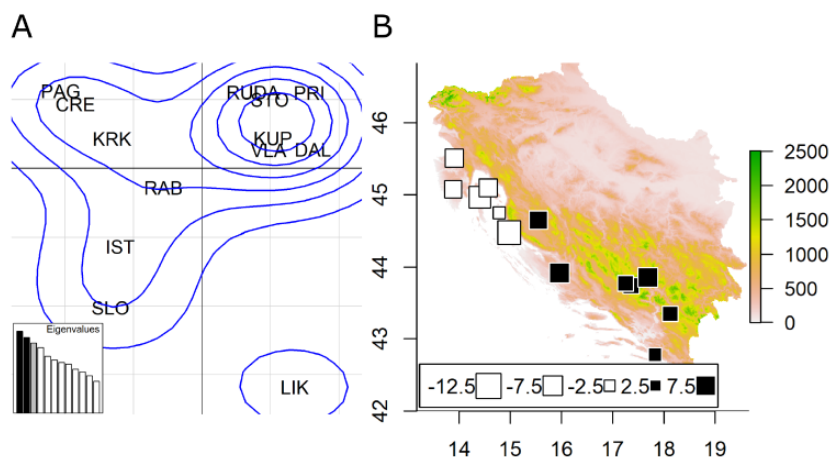
Sampled breed	Group abbreviation	Country of sampling	Climate	Longitude	Latitude
Cres Island Sheep	CRE	Croatia	Subtropical	44.96732	14.40567
Dalmatian Pramenka	DAL	Croatia	Subtropical	43.92263	15.95975
Istrian Sheep	IST	Croatia	Subtropical	45.07881	13.88324
Krk Island Sheep	KRK	Croatia	Subtropical	45.09899	14.56650
Kupres Pramenka	KUP	Bosnia and Herzegovina	Highland	43.74753	17.36175
Lika Pramenka	LIK	Croatia	Highland	44.64939	15.54746
Pag Island Sheep	PAG	Croatia	Subtropical	44.48131	14.97608
Privor Pramenka	PRI	Bosnia and Herzegovina	Continental	43.86151	17.68529
Rab Island Sheep	RAB	Croatia	Subtropical	44.75275	14.78397
Dubrovnik Ruda	RUDA	Croatia	Mediterranean	42.78214	17.82512
Istrian Pramenka	SLO	Slovenia	Subtropical	45.51118	13.90069
Stolac Pramenka	STO	Bosnia and Herzegovina	Subtropical	43.34809	18.12578
Vlašić Pramenka	VLA	Bosnia and Herzegovina	Continental	43.77771	17.25128

## Results and discussion

First three PCs (principal components) explained 34.62% of the genetic variability. Structure identified by PCA first three axes points to two patches, separating Lika Pramenka (LIK) as the most genetically different. Kernel density estimation suggests that LIK sheep belongs to a separate patch (Figure 1. A). Even though the first PC plotted onto the map exhibited two group of populations (from the northern part of the Eastern Adriatic, and from the Southern part), the pattern could not be identified as a cline or as two patches (Figure 1. B). As in the factorial correspondence analysis performed on the 12 Eastern Adriatic populations (43% of the variability explained in three components) LIK is the most distinguished population (Salamon et al., 2014). Second PC also separates the Pag Island Sheep (PAG), Cres Island Sheep (CRE) and Krk Island Sheep (KRK) from the Rab Island Sheep (RAB), Istrian Sheep (IST) from Croatia and Istrian Pramenka from Slovenia (SLO). The third axis contributes to separation of the SLO and IST populations from PAG, CRE, KRK, and RAB populations, as well as separation of Dubrovnik Ruda (RUDA) population from the Dalmatian Pramenka (DAL), Vlačić Pramenka (VLA), Kupres Pramenka (KUP), Stolac Pramenka (STO), Privor Pramenka (PRI) populations. The first three axes were tested for spatial autocorrelation. The Moran's I tests detected significant autocorrelation in the scores of the first PCA axis ( $I = 0.565$ ,  $p < 0.01$ ). The geographical component in the first PC was also reported by Laloë et al. (2010). The other two PCs did not show spatial autocorrelation ( $I = 0.040$ , NS;  $I = -0.065$ , NS).

underlying spatial processes were random, and more of the spatial content is explained by the first sPC (Figure 2. B). The first global scores component represents 18.4% of the total variation. According to Laloë et al. (2010), the first PC and sPC in sheep populations were highly correlated. The same was found in our results ( $r^2 = 0.845$ ,  $p < 0.001$ ). The negative scores (second sPC) did not exhibit a significant local structure according to local permutation test ( $\max(t) = 0.091$ , NS). Since the second sPC represents 22.3% of the total variation, it was used for the final interpretation (Figure 2. C. and D.). Lower correlation was found between the second sPC and the second PC ( $r^2 = 0.648$ ,  $p < 0.05$ ).

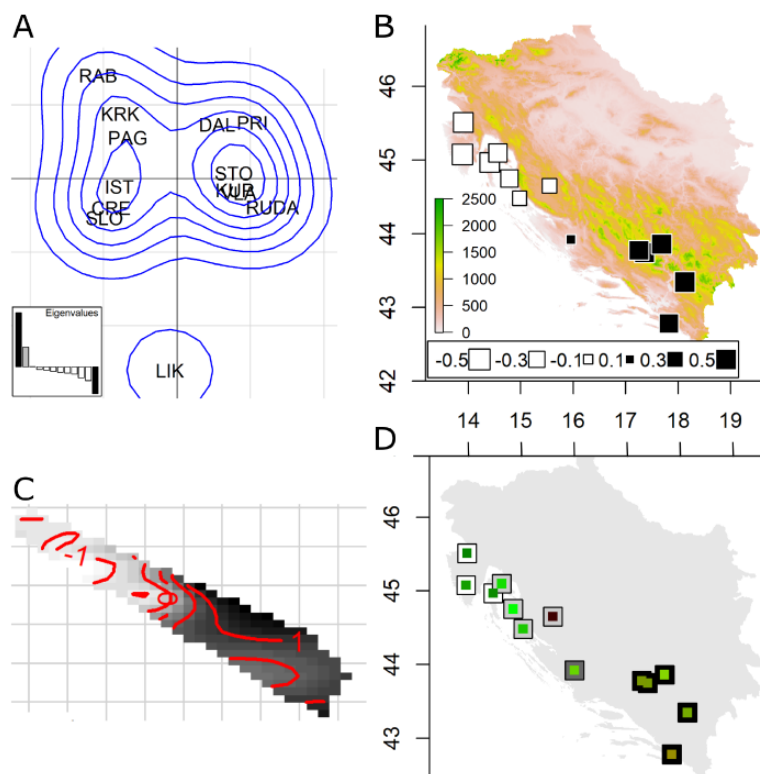
The global scores of the first sPC clearly differentiated the north-west (RAB, KRK, PAG, IST, CRE, SLO, LIK) from the south-east global structure (DAL, PRI, STO, KUP, VLA, RUDA) of populations (Figure 2. A). Kernel density estimation suggests this structure to be the larger patch showing substructure. The smaller patch suggested by the kernel density estimation contains only LIK population (Figure 2. A) and in sPCA this population is distinguished considering both sPCs. Lag vector of the first sPC plotted on the altitude map reveals the first global structure to be a cline spanning from southeast towards northwest (Figure 2. B). The southwest to northwest cline was found in other sheep populations, but also in cattle and goat populations (Laloë et al., 2010), and could be explained by SE-NW expansion after the domestication event. Local interpolation of scores points the steepest genetic differentiation south of Pag Island and Lika area where the red lines are concentrated, which is the area of DAL population (Figure 2. C). Similarly, model-based clustering in



**Figure 1.** Results of principal component analysis. A First and second principle axes shown with kernel density estimation to visualize structure in the data. B Representations of the first PC scores of the Eastern Adriatic sheep populations on the altitude map of Slovenia, Croatia and Bosnia and Herzegovina.

The first two spatial principle components (sPCs) explained 40% of the total variation, more than the tree PCs, and more than first five axes of the sPCA performed by Laloë et al. (2010) on microsatellite data of 46 sheep breeds. Global permutation test rejected the null-hypothesis of the absence of spatial structure ( $\max(t) = 0.110$ ,  $p < 0.05$ ). Higher level of the spatial content was found in the first sPC ( $I = 0.884$ ) than in the first PC ( $I = 0.565$ ), indicating the spatial distribution of high scores in the dataset is more spatially clustered than would be expected if

previous research proposed low cluster differentiation in DAL (Salamon et al., 2014). Second sPC divides the southern group DAL and PRI from STO, KUP, VLA and RUDA populations. It also divides LIK, IST, CRE and SLO populations from RAB, KRK and PAG populations in the northern group. This could be the reason why previous attempts with the hierarchical analysis of molecular variance did not detect pronounced differentiation between the mainland and the island breeds (Salamon et al., 2014). The final result of the sPCA shown on Figure 2. D



**Figure 2.** Results of spatial principle component analysis. A First positive and first negative scores component shown, with kernel density estimation to visualize structure in the data. B Lag vector of the first principle component plotted on the altitude map of Slovenia Croatia and Bosnia and Herzegovina. Sizes of squares represent different absolute values. C Spatial interpolation of scores in grey levels with contour lines in red indicating the genetic differentiation. D The results of spatial principle component analysis showing first global scores component and first negative scores component assigned to red and green colour channel, presented in geographical space.

presents the south-east to north-west cline and additional genetic differentiation captured from the second sPC in red-green-blue channel, emphasizing substructure of the northern part of the cline where SLO, IST and CRE are differentiated from KRK, RAB and PAG, and more pronouncedly from LIK population. In the southern part of the cline DAL and PRI are distinguished from STO, VLA, KUP and RUDA, with the last one showing more pronounced genetic differentiation inside its' group when second sPC is considered.

Despite the weak spatial structuring in sheep populations (Laloë et al., 2010) in comparison with other farm animals of domestic species, spatial aspect shows certain value in multivariate discovery-based studies and exploits the whole diversity in the data. Since model based clustering methods are useful in explaining diversity of the breeds as the management units (FAO, 2011), but are not recommended strategy when there is a cline structure behind the variability (Jombart et al., 2008), spatial multivariate methods should be used to understand the variability, especially if the breed is not a distinct genetic population as in Eastern Adriatic or in Baltic sheep breeds (Tapio et al. 2005).

## Conclusion

Focusing the analysis on the part of the variance that is spatially structured and using spatial information as a component of the optimized criterion in sPCA we have successfully retrieved south-east to north-west cline as the simple global structure, and even though the negative component did not have relevant spatial information, it provided explanation of the additional captured

variation, and performed better than the PCA. Likewise, sPCA explained the variation better than other multivariate methods (factorial correspondence analysis) and model based structuring methods (structure and AMOVA). Further investigation encompassing more breeds could reveal if the pronounced genetic differences found in LIK and RUDA sheep belong to certain spatial structure in general sheep population.

## References

- Bivand R., Piras G. (2015). Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw* 63(18): 1-36
- Bivand R., Lewin-Koh N. (2017). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.9-1. <https://CRAN.R-project.org/package=maptools>
- Duruz S., Flury C., Matasci G., Joerin F., Widmer I., Joost S. (2017). A WebGIS platform for the monitoring of Farm Animal Genetic Resources (GENMON). *PLoS One* 12(4): e0176362
- Food and Agriculture Organization of the United Nations (FAO) 2011. Molecular genetic characterization of animal genetic resources. FAO Animal production and health guidelines. No. 9. Commission on genetic resources for food and agriculture. FAO, Rome, Italy.
- François O., Waits L. P. (2016). Clustering and assignment methods in landscape genetics. In: *Landscape Genetics: Concepts, Methods, Applications*, First Edition. (N Balkenhol, S A Cushman, A T Storfer, L P Waits, eds), John Wiley & Sons, Ltd., 114-128
- Hijmans R. J. (2016). *raster: Geographic Data Analysis and Modeling*. R package version 2.5-8. <https://CRAN.R-project.org/package=raster>

- Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* (24): 1403-1405
- Jombart T., Devillard S., Dufour A. B., Pontier D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* (101): 92-103
- Joost S., Bruford M. W., The Genomic-Resources Consortium (2015). Editorial: Advances in farm animal genomic resources. *Front Genet* (6): 1-4
- Joost S., Colli L., Baret P. V., Garcia J. F., Boettcher P. J., Tixier-Boichard M., Ajmone-Marsan P., GLOBALDIV Consortium (2010). Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Anim. Genet.* 41(Suppl 1): 47-63
- Laloë D., Moazami-Goudarzi K., Lenstra J. A., Ajmone Marsan P., Azor P., Baumung R., Bradley D. G., Bruford M. W., Cañón J., Dolf G., Dunner S., Erhardt G., Hewitt G., Kantanen J., Obexer-Ruff G., Olsaker I., Rodellar C., Valentini A., Wiener P., European Cattle Genetic Diversity Consortium and Econogene Consortium (2010). Spatial trends of genetic variation of domestic ruminants in Europe. *Diversity*. 2: 932-945
- Legendre P., Legendre L (1998). Spatial analysis. In: Numerical ecology, Developments in Environmental Modelling. Second English Edition. (P Legendre, L Legendre, eds) Elsevier Science, Amsterdam, 736-738
- Manel S., Schwartz M. K., Luikart G., Taberlet P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol.* 18: 4 189-197
- Manni F., Guerard E., Heyer E. (2004): Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by “Monmonier’s algorithm”. *Hum Biol.* 76: 173–190
- Pariset L., Joost S., Gargani M., Valentini A. (2012). Landscape Genomics in Livestock. In: Analysis of genetic variation in animals (M Caliskan eds), InTech, pp 189-216
- Patterson N, Price A. L, Reich D (2006). Population structure and eigenanalysis. *PLOS Genetics*. 2(12): e190
- R Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ruiz-Garcia M., Jordana J. (2000). Spatial Structure and Gene Flow from Biochemical Markers in the “Pyrenean Brown” Breed, a Rare Cattle Race in Catalonia (Spain). *Biochem Genet.* 38: 341-352
- Salamon D., Gutierrez-Gil B., Arranz J. J., Barreta J., Batinic V., Dzidic A. (2014). Genetic diversity and differentiation of 12 eastern Adriatic and western Dinaric native sheep breeds using microsatellites. *Animal*. 8(2): 200–207
- Smouse P. E., Peakall R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*. 82: 561-573
- Šalamon D., Gutierrez-Gil B., Simčič M., Kompan D., Džidić A. (2015). Microsatellite based genetic structure of regional transboundary Istrian sheep breed populations in Croatia and Slovenia. *Mljekarstvo*. 65(1): 39-47
- Tapio I, Tapio M, Grislis Z, Holm L E, Jeppsson S, Kantanen J, Miceikiene I, Olsaker I, Viinalass H, Eythórsdóttir E (2005) Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity*. 94(4): 448-456