ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 68 6 Number 1, 2020

THE GENETIC STRUCTURE OF SLOVAK SPOTTED CATTLE BASED ON GENOME-WIDE ANALYSIS

Kristína Lehocká[1], Barbora Olšanská[1], Radovan Kasarda[1], Ondrej Kadlečík[1], Anna Trakovická[1], Nina Moravčíková[1]

[1] Department of Animal Genetics and Breeding Biology, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. Andreja Hlinku 609/2, 949 76 Nitra, Slovakia

Link to this article: https://doi.org/10.11118/actaun202068010057
Received: 25. 11. 2019, Accepted: 24. 1. 2020

To cite this article: LEHOCKÁ KRISTÍNA, OLŠANSKÁ BARBORA, KASARDA RADOVAN, KADLEČÍK ONDREJ, TRAKOVICKÁ ANNA, MORAVČÍKOVÁ NINA. 2020. The Genetic Structure of Slovak Spotted Cattle Based on Genome-wide Analysis. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 68(1): 57–61.

Abstract

The objective of the study was to determine the membership probability and level of admixture among Slovak Spotted cattle and historically related breeds (Ayshire, Holstein, Swiss Simmental and Slovak Pinzgau). The analysis was based on the panel of 35 934 SNPs that were used for genotyping of 423 individuals. The optimal number of clusters was estimated in two ways; by analysis of Bayesian information criterion and Bayesian clustering algorithm. The optimal number of clusters ranged from 3 to 5, depending on the applied approach. Subsequently, the population structure was tested by discriminant analysis of principal components (DAPC) and unsupervised Bayesian analysis based on the correlated allele frequencies model. The first discriminant function revealed three genetic clusters in population resulting from the production type and origin of analysed breeds. The unsupervised Bayesian analysis showed similar results, where the highest level of admixture was found between Slovak Pinzgau and Slovak Spotted cattle (0.6%). Despite that, the results of this study clearly showed that the Slovak Spotted cattle is genetically separated from other breeds that were involved in its grading-up process.

Keywords: dual-purpose breed, DAPC analysis, membership probability, population structure

INTRODUCTION

The detection and evaluation of population structure are one of the basic methods of modern population genetics. The populations' genetic structure can be estimated by different methods which use pedigree data or molecular markers information. The principal component analysis (PCA) belongs to the oldest and most widely technique of multivariate data analysis (Deperi *et al.*, 2018). According to Price *et al.* (2006) the PCA is frequently used tool in statistical genetics to derive the population structure from genomic data mostly based on genotyping of SNP (single nucleotide polymorphism) markers. Kadlečík *et al.* (2017) and Jombart *et al.* (2010) reported a summary of the overall variability among individuals as one of the objectives of PCA analysis. This summary includes divergence between groups and genetic variations occurring randomly in groups. Therefore, a different method is needed to evaluate the relationships between populations (clusters) focusing on the variability between clusters, which is the aim of discriminant analysis (DA) (Fisher, 1936; Jombart *et al.*, 2010).

The discriminant analysis is the multivariate method which serves to summarize the genetic differentiation and variation among groups. DA works on a similar principle as Bayesian clustering methods, permits for a probabilistic allotment of individuals to each group (Jombart *et al.*, 2010). However, the application of discrimination analysis on the genetic dataset is partially limited (Reyment, 2005). The new methodological approach that respects all limitations of DA without being hampered by its constraints is discriminant analysis of principal components (DAPC). The DAPC analysis allows to identify regions of the genome driving genetic divergence among clusters even if analysed groups are genetically related. The DAPC analysis implemented in R package Adegenet provides information about the visual evaluation of genetic structures among population (Jombart *et al.*, 2010). DAPC analysis similarly as Bayesian algorithm adopted in software STRUCTURE (Pritchard *et al.*, 2000) offers information about membership probabilities for each individual to each identified group. The Bayesian clustering algorithm is a tool for inferring population structure in population genetics which exploit genetic information to determine the membership of individuals in a population without predefined populations (Chen *et al.*, 2007).

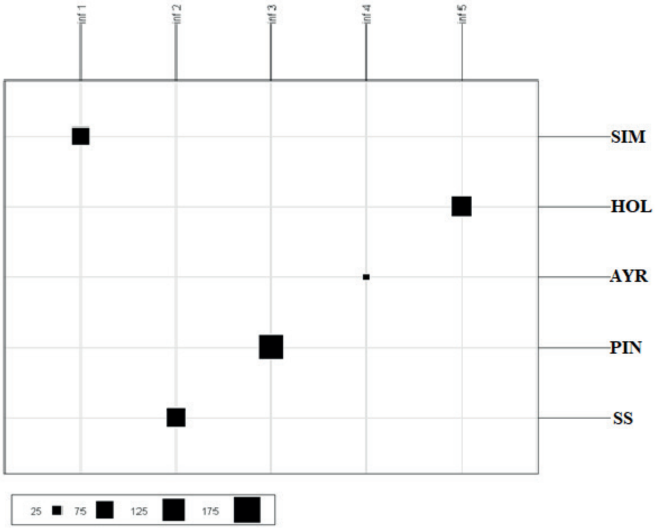
Aim of this study was to determine the genetic structure of the Slovak Spotted cattle in comparison to the historically relates breeds based on the medium-density SNP panel.

MATERIALS AND METHODS

SNP dataset was created by joining new and previously published data. In this study, we used genotypes of 35 934 autosomal SNP markers to characterize population structure. The dataset consisted of 423 animals representing Slovak Spotted cattle (N = 85) and historically related breeds; Holstein (N = 99), Swiss Simmental (N = 78), Slovak Pinzgau (N = 151) and Ayrshire (N = 10). The Slovak Spotted and Slovak Pinzgau cattle were genotyped by two platforms: Illumina BovineSNP50v2 BeadChip and ICBF International Dairy and Beef v3. The genotyping data of Holstein, Swiss Simmental and Ayrshire were obtained from web-based data repository (McTavish *et al.*, 2013). The quality control of SNP data was performed by program PLINK v1.9 (Chang *et al.*, 2015). According to the latest bovine genome assembly (Btau 5.0.1), all SNPs with an unknown chromosomal position and or located on sex chromosomes were excluded from database (Moravčíková *et al.*, 2018). Subsequently, the animals and SNPs with a call rate less than 0.90 and minor allele frequency (MAF) of 0.05 were excluded from the dataset. The final dataset included 34 817 SNP markers.

The genetic structure among the analysed breeds was estimated by two approaches, the Bayesian clustering method and DAPC analysis. The DAPC analysis was used to express relationships and genetic structure among populations by using the R package Adegenet (Jombart and Collins, 2015). The optimal number of genetic clusters was determined by running the sequential K-means algorithm that uses the Bayesian Information Criterion (BIC). The model was used to search the likelihood associated with each value of the K-means clustering algorithm from K = 1 to K = 40. We used DAPC analysis to assign individuals and determine the membership probability that represents the total genetic background of an individual. The 100 PC and 10 discriminant functions were run used in the DAPC analysis.

The Bayesian statistics adopted in program Structure 2.3.4 (Pritchard *et al.*, 2000) was then used for the estimation of population structure. The analysis was performed using the correlated allele frequencies model across all individuals using the burn-in period of 1 000 permutations



1: *The BIC statistic referred to the differentiation between inferred (inf) and original clusters among evaluated breeds (SIM – Swiss Simmental, HOL – Holstein, AYR – Ayrshire, PIN – Slovak Pinzgau, SS – Slovak Spotted.)*

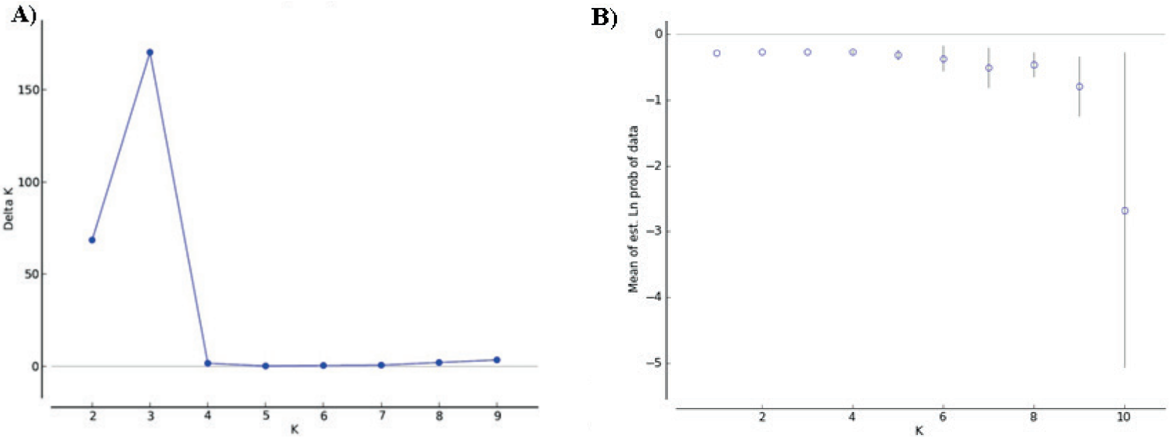
followed by 10 000 MCMC replications. The runs were performed from K = 1 to K = 10. According to Evanno *et al.* (2005) the optimal value of clusters were analysed based on the logarithmic delta K (ΔK) probability assessment by program STRUCTURE HARVESTER v0.3 (Earl and von Holdt, 2012).

RESULT AND DISCUSSION

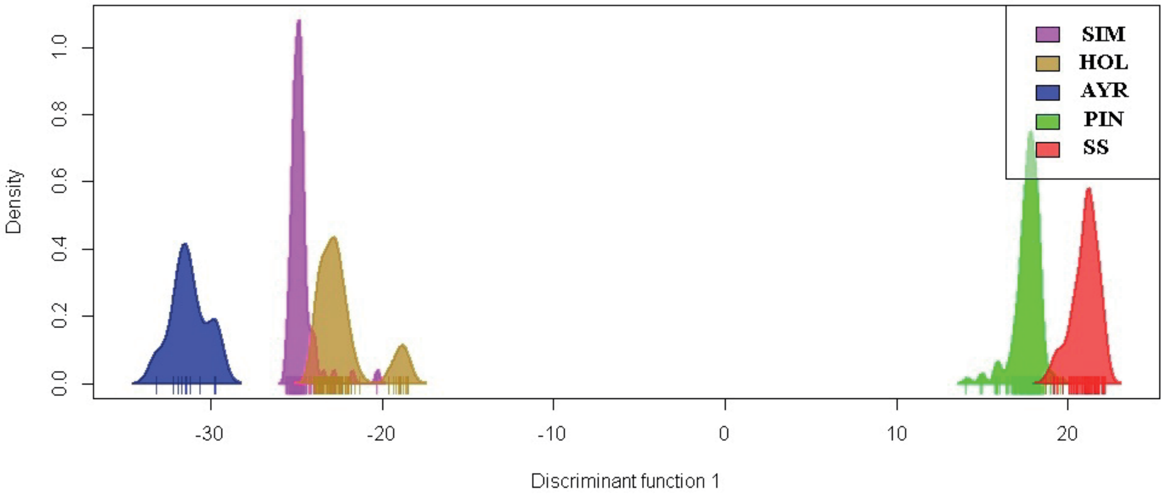
The optimal number of clusters was estimated by two approaches; analysis of Bayesian information criterion (BIC) and Bayesian clustering algorithm. The BIC analysis showed that the division of individuals inferred clusters corresponded to the initial groups and the optimal value was K = 5 (Fig. 1). The method according to Evanno *et al.* (2005) that analyse the differences of change of the likelihood with respect to ΔK showed as the optimal number of clusters K = 3 (Fig. 2A). However, the likelihood of the observed database considering the number of the expected population of origin [Ln Pr(X | K)] revealed as optimal K = 5 (Fig. 2B).

The DAPC analysis showed certain level of genetic differentiation among historically related breeds. The results indicated that some of the analysed breeds were clustered together probably due to common historical origin. The first discrimination function revealed formation of three clusters; first formed from Ayrshire, second from Holstein and Swiss Simmental, and third composed from Slovak Spotted and Slovak Pinzgau cattle (Fig. 3). Based on the first and second discriminant function we were able to describe 62% of variability in dataset. As shown in Fig. 4, the origin of analysed breeds resulted in formation of three clusters, but each breeds can be regarded as genetically separated. The close genetic affinity between Slovak Spotted and Slovak Pinzgau cattle was previously reported by Židek *et al.* (2008).

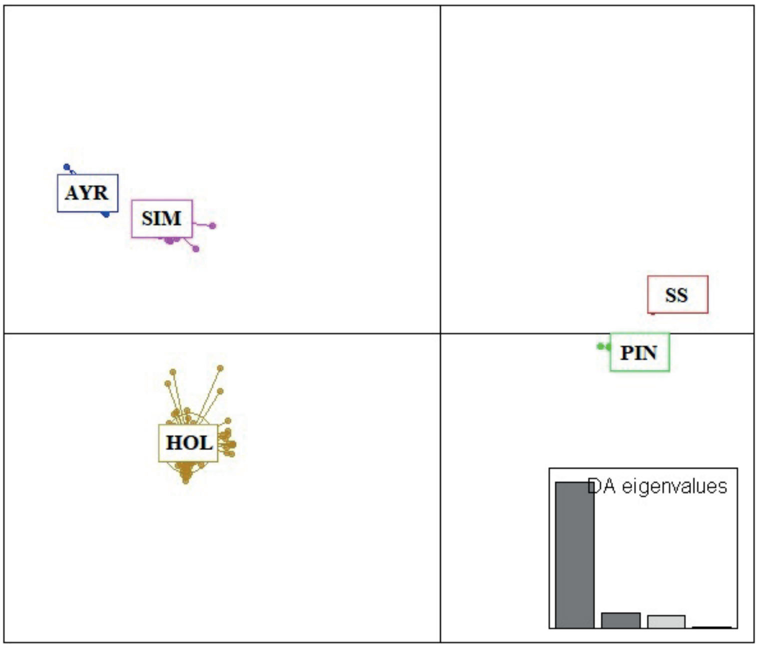
To confirm the results from DAPC analysis, the proportion of admixture among breeds was tested by the Bayesian clustering algorithm. Tab. I summarizes the membership probability resulted from STRUCTURE analysis. The first cluster corresponded to the Holstein cattle with



2: *Delta K calculated by method proposed by Evanno et al. (2005) (A) and mean of estimated likelihood L(K) and variance per K values (B)*



3: *The first discriminant function determined the genetic clusters*



4: *The first two discriminant functions determined the genetic clusters*

I: *The proportion of the membership probability based on the Bayesian clustering approach*

Predefined populations	Inferred clusters					Number of individual
	1	2	3	4	5	
SIM	0.067	0.043	0.049	0.176	**0.84**	78
HOL	**0.914**	0.019	0.004	0.055	0.01	99
AYR	0.133	0.162	0.002	**0.687**	0.02	10
PIN	0.014	**0.551**	0.042	0.004	0.002	151
SS	0.0093	0.034	**0.944**	0.006	0.01	85

membership probability higher than 91%. The second cluster is composed from Slovak Pinzgau cattle (55%), but the membership probability showed that this population is affected by admixture. The third cluster corresponded to Slovak Spotted cattle (94%), with only negligible proportion of admixture. The fourth cluster included Ayrshire breed with membership probability at level 68%. The Swiss Simmental breed was assigned in the fifth cluster with membership probability up to 84%.

With respect to the genome of Slovak Spotted cattle the highest proportion of admixture was found for Slovak Pinzgau (0.6%) and Swiss Simmental cattle (1%). The obtained results reflected the grading-up process of Slovak Spotted cattle. From a historical point of view, the Slovak Spotted and the Slovak Pinzgau cattle have consisted of extinct autochthonous Carpathian Red and Carpathian Grey breeds. In the 18th and 19th centuries, Swiss Simmental and Austrian Pinzgau were involved in breeding on the Slovak dual-purpose breeds (Kasarda *et al.*, 2015). After 1972, the Ayrshire and Holstein bulls were used in the selection process to increase production.

CONCLUSION

In conclusion, this study provides insight into the genetic structure of Slovak Spotted cattle, which belongs to one of two autochthonous breeds kept in Slovakia. Slovak Spotted cattle based on our results can be referred to as a relative genetic distinct breed. The results clearly showed a high level of admixture among Slovak Spotted and among evaluated breeds. The highest values of admixture were achieved in Slovak Pinzgau and Swiss Simmental breeds due to the grading-up process in history. The results will provide important information that can be used for the conservation of genetic diversity as well as in the development of sustainable breeding strategies.

Acknowledgements

The Slovak Research and Development Agency (APVV-14-0054 and APVV-17-0060) and VEGA (1/0742/17) this study was supported.

REFERENCES

CHANG, C. C. *et al.* 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1): 7.

CHEN, C. *et al.* 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes,* 7(5): 747–756.

DEPERI, S. I. *et al.* 2018. Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLoS ONE*, 13(3): e0194398.

EARL, D. A. and VON HOLDT, B. M. 2012. Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res*., 4: 359–361.

EVANNO, G., REGNAUT, S. and GOUDET, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol*., 14(8): 2611–2620.

FISHER, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2): 179–188.

JOMBART, T., DEVILLARD, S. and BALLOUX, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetic*, 11: 94.

JOMBART, T. and COLLINS, C. 2015. *A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0*. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.

KADLEČÍK, O., MORAVČÍKOVÁ, N. and KASARDA, R. 2017. *Biodiversity of animal populations* [in Slovak: *Biodiverzita populácií zvierat*]. Nitra: SUA University Press.

KASARDA, R. *et al.* 2015. Genome-wide selection signatures in pinzgau cattle. *Potravinarstvo*, 9(1): 268–274. [Online].

MCTAVISH, E. J. *et al.* 2013. New world cattle show ancestry from multiple independent domestication events. *Proc. Natl. Acad. Sci. USA*, 110(15): 1398–1406.

MORAVČÍKOVÁ, N. *et al.* 2018. Autozygosity island resulting from artificial selection in slovak spotted cattle. *Agriculture & Forestry*, 64(4): 21–28.

PRICE, A. L. *et al.* 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, 3: 904–909.

PRITCHARD, J. K., STEPHENS, M. and DONNELLY, P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2): 945–959.

REYMENT, R. A. 2005. The statistical analysis of multivariate serological frequency data. *Bulletin of Mathematical Biology*, 67(6): 1303–1313.

ŽIDEK, R. *et al.* 2008. comparison of microsatellite and blood group diversity among different genotypes of cattle. *Acta Veterinaria Hungarica*, 56(3): 323–333.

Contact information

Kristína Lehocká: tina.lehocka@gmail.com
Barbora Olšanská: baska.olsanska@outlook.com
Radovan Kasarda: radovan.kasarda@gmail.com
Ondrej Kadlečík: ondrej.kadlecik@uniag.sk
Anna Trakovická: anna.trakovicka@uniag.sk
Nina Moravčíková: nina.moravcikova1@gmail.com