Original Scientific paper
10.7251/AGRENG2002090O
UDC 575.1:636.2

DETECTION OF SELECTION SIGNALS IN CATTLE POPULATIONS BY PCA

Barbora OLŠANSKÁ*, Radovan KASARDA, Kristína LEHOCKÁ, Nina MORAVČÍKOVÁ

Slovak University of Agriculture, Department of Genetics and Animal Breeding Biology, Trieda A. Hlinku 2, 949 76, Nitra, Slovakia
*Corresponding author: baska.olsanska@outlook.com

ABSTRACT

The presented study provides a genome-wide scan of selection signals in cattle by principal component analysis (*PCA*). The aim was to identify SNP affected by intensive selection based on package *PCAdapt* implemented under software *R*. This analysis provided insight into the association between the SNP frequencies related to population differentiation. The four cattle populations were involved in the analysis (Slovak Spotted cattle, Ayrshire, Swiss Simmental and Holstein) with overall 272 of genotyped individuals. After applying quality control, the final dataset consisted of 35 675 SNPs, with an overall length of 2496.14 Mb and average space between adjacent SNP 70.03 ± 76.1 kb. After performing *PCA* analysis, the uniqueness of the breeds was revealed. On the other hand, a close genetic relationship and eleven SNPs affected by selection were found, with a position close to 162 genes involved in the various biological processes. The majority of genes were involved in the positive regulation of adenylate cyclase activity, embryo development and somatic diversification of immune receptors via somatic mutation. Several candidate genes for genetic control of the immune system (*DNAJB9*), muscle development (*SEPT7, TRIM32, ROCK1, NRAP, PZDZ8, HSPA12A* and *FGFR2*), milk production (*SOCS5, CD46*), reproduction *(LHCGR, EEPD1, FSHR*) and coat colour (*KIT*) were identified. Our results provide insights into the regions of the genome affected by the intensive selection of analysed cattle populations.

Keywords: *biological process, Bos Taurus, footprints of selection, PCAdapt, production traits.*

INTRODUCTION

Nowadays, cattle are one of the most important livestock species in the world. Cattle is significant because humanity benefits through their production of milk, meat, leather, and traction force. Since the domestication process, directional selection has caused different phenotypic and biological characteristics of different cattle breeds (Randhawa et al., 2016). In Slovakia, cattle breeding is part of a

closed agricultural system and a vital co-creator of the environment. Non-productive functions of cattle breeding significantly contribute to maintaining the cultural character of the country and its social function (Brestenský et al., 2015).
Bovine breeds that are considered local are a source of significant genetic variation because they may have alleles that allow them to adapt to the local environment and feed (Bahati et al., 2020). Directional selection has been focused on achieving the breeding objectives provided the individual's adaptation to the specific production environment (Duforet-Frebourg et al., 2015). Such an impact of selection has left distinct genomic signatures that can be used to retrospectively understand breeding objectives, detect economically important alleles, and potentially allow for a better understanding of biologically specific phenotypes (Joukharad et al., 2019; Moradian et al., 2020).
For identification of the selection impact on genome architecture and detection of particular candidate genomic regions affected by intense selection, various analytical methods have been developed. Previous studies in cattle provide information about selection which affected genome variation and architecture. Selection for individual breeds led to the fixation of specific variants, which in comparison with others became specific signatures for each breed. Such selection signals were observed in different populations using several methods, e.g. based on the allelic frequency spectrum, differentiation of populations (F_{ST}) an extent of linkage disequilibrium (Mustafa et al., 2018; Cheruiyot et al., 2018; Moravčíková et al., 2017). Using principal component analysis (PCA), whole-genome selection scans can be performed. The standard F_{ST} index of genetic differentiation between populations can be considered as the variance ratio explained by the major principal components (Duforet-Frebourg et al., 2015).
The objectives of this study were to describe population structure, identify footprints of selection based on PCA analysis as well as genes in particular genomic regions and their function in the biological processes in cattle with a focus on the genome of Slovak Spotted, Swiss Simmental, Holstein and Ayrshire breeds.

MATERIAL AND METHODS

For this analysis, 272 genotyped animals belonging to four cattle breeds (Slovak Spotted - SS, Swiss Simmental - SIM, Holstein – HOL, Ayrshire – AYR) were used. The final dataset included data for 85 SS animals genotyped by the Illumina BovineSNP50v2 BeadChip (bulls) and ICBF International Dairy and Beef v3 (dams). The genotypes of other breeds: 78 SIM, 99 HOL and 10 AYR described in McTavish et al. (2013) were used. After data merging, quality control was done according to Moravčíková et al. (2017).
Outlier loci signalising selection signals in particular genomic regions were identified based on the principal component analysis (PCA) by package *PCAdapt* (Luu et al., 2020). Genetic differentiation within and across analysed breeds was tested based on the Mahalanobis distance approach. The Mahalanobis distance is a multi-dimensional method to find out the distance of a point from the mean based

on the *Z*-scores gained from the regression of SNPs with K principal components. The Mahalanobis distance is defined as:

$$D_j^2 = \left(z_j - \bar{z}\right)^T \Sigma^{-1} (z^j - \bar{z})$$

where Σ is the (K × K) covariance matrix of the z-scores and $\bar{z}$ is the vector of the K z-score means. The Mahalanobis distance had to be transformed to P-values to achieve a real value of number between 0 and 1. The P-values less than 0.05 was considered as outliers signalising significant impact of selection on the particular region in the autosomal genome (Luu et al., 2020). For visualisation of p-values was used package *qqman* under R software environment (Turner 2017). After targeted the outlier loci, genes involved in various biological processes (Genome data viewer, WebGestalt) and QTLs (CattleQTLdb) were identified.

RESULTS AND DISCUSSION

The final dataset consisted of 35 675 SNP with average space between adjacent SNP 70.03 ± 76.01 kb and the overall length of the genome was 2496.14 Mb. Using *PCA* method, the genetic structure of the analysed populations was defined and a whole-genome scan of SNPs associated with selection was performed. Obtained individual membership probabilities indicated grouping of animals into the genetic clusters concerning specific breeding objectives of each breed (Figure 1). The Holstein and Swiss Simmental populations created separate genetic clusters. Figure 1 indicated that Slovak Spotted and Ayrshire populations were partially linked together probably due to use of Ayrshire in the grading-up of Slovak Spotted.

Huson et al. (2020) reported for Jersey cattle bred in different areas homogeneity despite a various geographic origin, compared with Holstein and Guernsey. In contrary, Ahmad et al. (2020) found among breeds Holstein, Jersey and Brown Swiss high degree of differentiation depending on the geographical origin. Cheruiyot et al. (2018) observed dispersion of Tanzanian crossbred cattle toward Northern European taurine breeds like Holstein, Norwegian Red and Friesian. These results confirm our suggestions that close genetic relationships among analysed breeds arise most likely due to cross-breeding for improvement of Slovak Spotted performance traits. The subsequent analyses of selection signatures showed that the artificial selection acted mainly in the genomic regions associated with milk and meat production, reproduction and exterior characteristics. Signatures of intensive selection in the genome are highlight in Table 1.

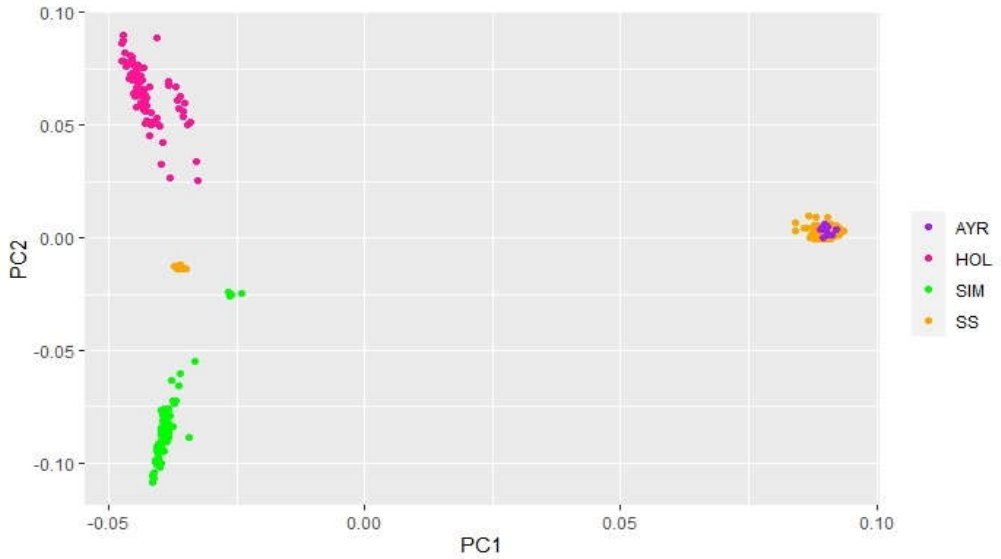

Figure 1. Structure of analysed populations used to determine SNPs associated with selection (Slovak Spotted - SS, Swiss Simmental - SIM, Holstein – HOL, Ayrshire – AYR),

Table 1. Description of 10 regions affected by intense selection

CHR	Position (Mb)	QTL
4	56.48 - 63.12	height (yearling), carpic acid content, clinical mastitis, milk palmitoleic content, parasites mean of natural logarithm, milk protein percentage, muscle carnosine content, intramuscular fat
6	69.88 - 80.89	milk yield, milk protein yield, conception rate, milk fat yield
8	101.50 - 110.65	milk fat percentage, milk protein percentage, milk butyric acid content, milk caproic acid content, milk caprylic acid content, milk capric acid content, milk lauric acid content, milk myristic acid content, milk palmitic acid content, milk margaric acid content
11	23.72 - 32.90	shear force, marbling score, milk palmitic, lean meat yield
13	68.63 -71.01	teat length, first service conception, conception rate, milk caproic acid content, milk caprylic acid content, milk capric acid content, milk myristoleic acid content, milk palmitoleic acid content
16	70.99 - 79.58	milk kappa-casein percentage, first service conception, conception rate, conformation score, social separation - standing alert, milk protein percentage, muscle phosphorus content, interval to first estrus after calving
22	43.77 - 45.49	milk yield, conception rate, dairy form, PTA type, teat placement - front, udder attachment, udder height, rump width, udder cleft

23	30.16 - 30.71	milk protein yield, milk conjugated linoleic acid content, milk protein percentage, milk yield, milk mid-infrared spectra, immunoglobulin G level
24	29.41 - 37.19	cooking loss, oleic acid content, stillbirth, body weight (yearling), fat thickness at the 12th rib, body temperature, respiratory rate, milk tridecylic acid content, bovine respiratory disease susceptibility, udder depth
26	35.59 - 49.14	milk protein yield, milk fat yield, udder width, milk kappa-casein percentage, milk glycosylated kappa-casein percentage, conformation score

Manhattan plot (Figure 2) showed the SNP outliers identified using a whole-genome scan for selection signatures by principal component analysis. In the affected regions, 162 genes affected immune system, muscle development, milk production, reproduction and coat colour were identified. Fontanesi et al. (2010) reported that gene *KIT* might affect the coat colour. Gene *DNAJB9* was associated with inflammatory, immune and stress response in lambs (Sabino et al., 2018). Li et al. (2020) observed that gene *EEPD1* was associated with meat quality and reproduction traits in Holstein bulls. Saatchi et al. (2014) find out that gene *SEPT7* was a candidate gene for QTL affected body weights in Maine-Anjou cattle. Bongiorni et al. (2016) observed that gene *TRIM32* play a role in regulating skeletal muscle differentiation and regeneration of adult skeletal muscle in Maremmana and Chianina cattle. Gene *SOCS5* was distinctly expressed during the lactation in Australian dairy breed (Arun et al., 2015). Wohlres-Viana et al. (2017) reported that gene *LHCGR* control LH effect during follicle growth and ovulation. Gene *FSHR* was associated with promoting follicle of Yaks (Xia et al., 2020). Wang et al. (2014) suggested that polymorphism of gene *CD46* is associated with the occurrence of mastitis in dairy cow. Gene *NR5A2* is included in the processes of spermatogenesis in mice (Liu et al., 2016). Taye et al. (2017) reported relatedness of gene *ROCK1* with intramuscular fat. Genes *NRAP, PZDZ8* and *HSPA12A* were associated with a muscular function (Masoudi et al., 2008). Akizawa et al. (2016) reported that gene *FGFR2* was involved in the regulation of inner cell mass development and blastocyst formation in cattle.

Compared to our results, Mustafa et al. (2018) reported for *Bos Indicus* 11 outlier loci with candidate genes associated with the immune system, muscle growth and some economically important traits like body growth and longevity. Cheruiyot et al. (2018) observed for dairy cattle in Tanzania candidate regions on chromosomes 6, 7, 14, 18 and 20. Flori et al. (2009) find selection signatures in 16 genomic regions for French dairy cattle. These findings indicated that the occurrence of selection signatures in genomic regions was caused by improve production characteristics or adapt to the environment. Genes involved in various biological processes (Table 2) have been identified in genomic regions under selection pressure. Observed biological processes suggested that the genome was affected by the artificial as well as natural selection to adapt specific production environment.

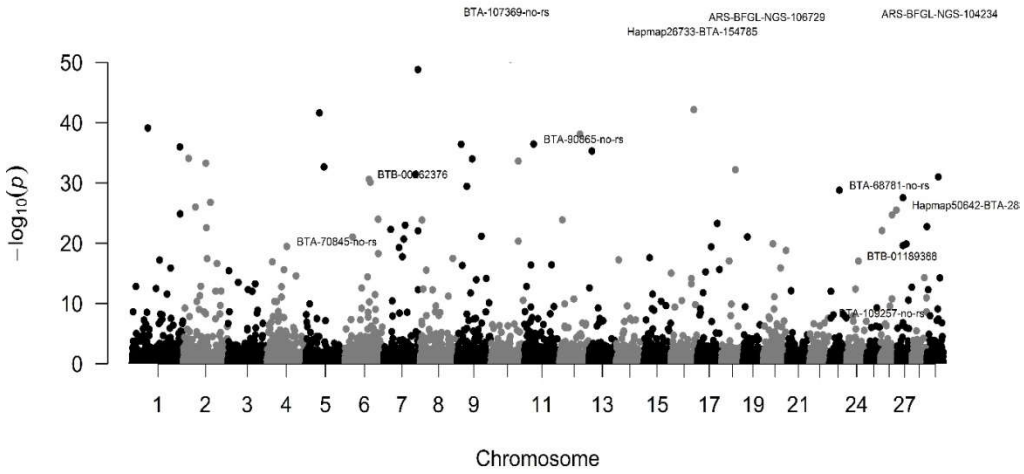

Figure 2. Manhattan plot of genomic regions indicating the SNPs associated with selection

Table 2. Biological processes of genes identified in genomic regions of outlier's loci in analysed populations

Gene Set	Description	p-value	Gene Code
GO:0051096	positive regulation of helicase activity	0.0005	*MSH2, MSH6*
GO:0051095	regulation of helicase activity	0.0007	*MSH2, MSH6*
GO:0045762	positive regulation of adenylate cyclase activity	0.0010	*FSHR, LHCGR*
GO:0009790	embryo development	0.0011	*HOPX, LAMB3, MIN1, MSH2, PLCG1, RBBP8, SEPT7, SIX3, TBX20, TOP1, WHRN*
GO:0051239	regulation of multicellular organismal process	0.0013	*AFAP1L2, BMPER, BRINP1, HOPX, KDR, LRRN3, MIB1, MSH2, MSH6, NMU, PLCG1, PPM1B, PRKCE, REST, SEPT7, SIX3, SOCS5, TBX20, TLR4, TRIM32*
GO:0007190	activation of adenylate cyclase activity	0.0013	*FSHR, LHCGR*
GO:0002566	somatic diversification of immune receptors via somatic mutation	0.0013	*MSH2, MSH6*
GO:0016446	somatic hypermutation of immunoglobulin genes	0.0013	*MSH2, MSH6*
GO:0071229	cellular response to acid	0.0020	*BRINP1, KDR, PRKCE, YES1*

	chemical		
GO:0048015	phosphatidylinositol-mediated signalling	0.0021	*EXOC1, FSHR, KCNH1, KDR*
GO:0048017	inositol lipid-mediated signalling	0.0023	*EXOC1, FSHR, KCNH1, KDR*
GO:0008340	determination of adult lifespan	0.0031	*MSH2, MSH6*
GO:0051240	positive regulation of multicellular organismal process	0.0036	*AFAP1L2, BMPER, BRINP1, KDR, LRRN3, MSH2, MSH6, NMU, PLCG1, PRKCE, SOCS5, TBX20, TRIM32*
GO:0045830	positive regulation of isotype switching	0.0036	*MSH2, MSH6*
GO:0045761	regulation of adenylate cyclase activity	0.0042	*FSHR, LHCGR*

CONCLUSION

The presented study confirmed that the genome of Slovak Spotted cattle was affected mainly by cross-breeding with Ayrshire cattle in the past to increase milk production. The genetic differences among analysed breeds were represented by ten genomic regions significantly affected by selection. The results showed that in the genome of analysed breeds specific alleles associated with adaptation to the production environment, performance traits and reproduction were fixed. The genes identified in the affected regions were involved in various processes, including regulation of immune system (*DNAJB9*), muscle development (*SEPT7, TRIM32, ROCK1, NRAP, PZDZ8, HSPA12A* and *FGFR2*), milk production (*SOCS5, CD46*), reproduction (*LHCGR, EEPD1, FSHR*) and coat colour (*KIT*). This study provided information about the breeding history of Slovak Spotted cattle and offer a basis for further studies to complement the knowledge about the associations between the genotype and phenotype of the Slovak Spotted cattle.

ACKNOWLEDGEMENT

This study was supported by the Slovak Research and Development Agency (APVV-14-0054 and APVV-17-0060).

REFERENCES

Ahmad, S.F., Panigrahi, M., Chhotaray, S., Pal, D., Parida, S., Bhushan, B.,Gaur, G.K., Mishra, B.P., Singhc, R.K. (2020). Revelation of genomic breed composition in a crossbred cattle of India with the help of Bovine50K BeadChip, Genomics, 112(2), (1531-1535).

Arun, S.J., Thomson, P.C., Sheehy, P.A., Khatkar, M.S., Raadsma, H.W., Williamson, P. (2015). Targeted Analysis Reveals an Important Role of JAK-STAT-SOCS Genes for Milk Production Traits in Australian Dairy Cattle, Front.Genet.

Bongiorni, S., Gruber, C.E.M., Bueno, S., Chillemi, G., Ferrè, F., Failla, S., Moioli, B., Valentini, A. (2016). Transcriptomic investigation of meat tenderness in two Italian cattle breeds, Animal Genetics, 47(3).

Brestenský, V. (2015) Chov hospodárskych zvierat, Národné poľnohospodárske a potravinárske centrum - Výskumný ústav živočíšnej výroby Nitra.

Cheruiyot, E.K., Bett, R.C., Amimo, J.O., Zhang, Y., Mrode, R., Mujibi, F.D.N. (2018). Signatures of Selection in Admixed Dairy Cattle in Tanzania, Front. Genet.

Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, R., Blum, M.G.B. (2015). Detecting Genomic Signatures of Natural Selection with Principal Component Analysis: Application to the 1000 Genomes Data, Molecular Biology and Evolution, 33(4), (1082-1093).

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., et al. (2009). The genome response to artificial selection: a case study in dairy cattle, PloSone, 4(8): e6595.

Fontanesi, L., Tazzoli, M., Russo, V., Beever, J. (2010). Genetic Heterogeneity at the Bovine KIT Gene in Cattle Breeds Carrying Different Putative Alleles at the Spotting Locus, Anim Genet., 41(3), (295-303).

Huson, H.J., Sonstegard, T.S., Godfrey, J., Hambrook, D., Wolfe, C., Wiggans, G., Blackburn, H., VanTassell, C.P. (2020). A Genetic Investigation of Island Jersey Cattle, the Foundation of the Jersey Breed: Comparing Population Structure and Selection to Guernsey, Holstein, and United States Jersey Cattle, Front. Genet.

Joukharad, R., Daetwyler, H.D., Gendall, A.R., Hayden, M.J. (2019). Artificial selection causes significant linkage disequilibrium among multiple unlinked genes in Australian wheat, Evolutionary Applications, 12(8).

Li, Y., Wang, M., Li, Q., Gao, Y., Li, Q., Li, J., Cao, Y. (2020). Transcriptome profiling of longissimus lumborum in Holstein bulls and steers with different beef qualities, PLoS ONE, 15(6): e0235218.

Liu, C., Qian, P., Yang, L., Zhang, L., Chen, C., He, M., et al. (2016). Pubertal exposure to di-(2-ethylhexyl)-phthalate inhibits G9a-mediated histone methylation during spermatogenesis in mice, Arch. Toxicol., 90, (955–969).

Luu, K., Blum, M., Privé, F., Bazin E., Duforet-frebourg, N. (2020) Fast Principal Component Analysis for Outlier Detection.

Masoudi, A.A., Uchida, K., Yokouchi, K., Ohwada, K., Abbasi, A.R., Tsuji, T., Watanabe, T., Hirano, T., Sugimoto, Y., Kunieda, T. (2008). Linkage mapping of the locus responsible for forelimb-girdle muscular anomaly of Japanese black cattle on bovine chromosome 26, Animal Genetics, 39(1).

McTavish E.J, Decker, J.E., Schnabel, R.D., Taylot, J.F., Jillis, D.M. (2013). New world cattle show ancestry from multiple independent domestication events, PNAS, 110(15): (1398–1406).

Moradian, H., Koshkoiyeh, A.E., Mohammadabadi,M., Fozi, M.A. (2020). Whole genome detection of recent selection signatures in Sarabi cattle: a unique Iranian taurine breed, Genes & Genomics, 42, (203-215).

Moravčíková, N., Kukučková, V., Mészáros, G., Sölkner, J., Kadlečík, O., Kasarda, R. (2017). Assessing footprints of natural selection through PCA analysis in cattle, Acta fytotechn zootechn, 20(2), (23-27).

Mustafa, H., Khan, W.A., Sonstegard, T., Li, Y., Ain, N.U., Ajmal, A. (2018). Genome-Wide Identification of Natural Selection Footprints in Bos Indicus Using Principal Component Analysis, Life Sci., 5(2), (67-72).

Randhawa, I.A.S., Khathar M.S., Thomson, P.C., Raadsma, H.W. (2016). A Meta-Assembly of Selection Signatures in Cattle, PLoS ONE 11(4): e0153013.

Saatchi, M., Schnabel, R.D., Taylor, J.F., Garrick, D.J. (2014). Large-effect Pleiotropic or Closely Linked QTL Segregate Within and Across Ten US Cattle Breeds, BMC Genomics, 15(1), (442).

Sabino, M., Carmelo, V.A.O., Mazzoni, G., Cappelli, K., Capomaccio, S., Ajmone-Marsan, P., Verini-Supplizi, A., Trabalza-Marinucci, M., Kadarmideen, H.N. (2018). Gene Co-Expression Networks in Liver and Muscle Transcriptome Reveal Sex-Specific Gene Expression in Lambs Fed With a Mix of Essential Oils, BMC Genomics, 19(1), (236).

Taye, M., Kim, J., Yoon, S.H., Lee., W., Hanotte, O., Dessie, T., Kemp, S., Mwai, O.A., Caetano-Anolles, K., Cho, S., Oh, S.J., Lee, H., Kim, H. (2017). Whole Genome Scan Reveals the Genetic Signature of African Ankole Cattle Breed and Potential for Higher Quality Beef, BMC Genet., 18(1), (11).

Turner, S. (2017). Q-Q and Manhattan Plots for GWAS Data.

Wang, X., Zhong, J., Yundong, G., Ju, Z., Huang, J. (2014). A SNP in Intron 8 of CD46 Causes a Novel Transcript Associated With Mastitis in Holsteins, BMC Genomics., 15(1), (630).

Wohlres-Viana, S., Arashiro, E.K.N., Minare, T.P., Fernandes, C.A.C., Grazia, J.G.V., Siqueira, L.G.B., Machado, M.A., Viana, J.H.M. (2017). Differential expression of LHCGR and its isoforms is associated to the variability in superovulation responses of Gir cattle, Theriogenology, 126, (68-74).

Xia, Y., Wang, Q., He, X.D., Chen, Y., JiGe, M.T., Zi, X.D. (2020). Cloning and Expression Analysis of the Follicle-Stimulating Hormone Receptor (FSHR) Gene in the Reproductive Axis of Female Yaks (Bos Grunniens), Domest Anim Endocrinol., 70, (106383).