https://doi.org/10.17221/265/2020-CJAS

The evaluation of genomic diversity and selection signals in the autochthonous Slovak Spotted cattle

Radovan Kasarda[1], Nina Moravčíková[1], Barbora Olšanská[1], Gábor Mészáros[2], Luboš Vostrý[3], Hana Vostrá-Vydrová[4], Kristína Lehocká[1], Ján Prišťák[1], Juraj Candrák[1]*

[1]Department of Animal Genetics and Breeding Biology, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic
[2]Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria
[3]Department of Genetics and Breeding, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic
[4]Department of Ethology and Companion Animal Science, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

**Corresponding author: nina.moravcikova@uniag.sk*

Citation: Kasarda R., Moravčiková N., Olšanská B., Mészáros G., Vostrý L., Vostrá-Vydrová H., Lehocká K., Prišťák J., Candrák J. (2021): The evaluation of genomic diversity and selection signals in the autochthonous Slovak Spotted cattle. Czech J. Anim. Sci., 66: 251–261.

Abstract: The aim of this study was to evaluate the effective population size based on linkage disequilibrium and the trend of inbreeding derived from runs of homozygosity (ROH) in the Slovak Spotted cattle. The ROH segments longer than 4 Mb were then analysed to identify selection signals. Eighty-five individuals were genotyped using the ICBF International Dairy and Beef chip (dams of sires) and Illumina BovineSNP50 BeadChip (sires). The ROH segments > 1 Mb occurred most often in the autosomal genome with the average number of 16.75 ± 7.23. The ROH segments > 16 Mb covering 0.41% of the genome pointed to the long-term effort of breeders to reduce inbreeding in the population of Slovak Spotted cattle. However, the average observed heterozygosity indicated a decrease in overall diversity in the current population. The decrease of heterozygosity per generation also confirmed the estimates of historical and recent effective population size (a decrease of 6.88 animals per generation). The predicted current effective population size was 58 animals. Twenty-one regions across 12 different autosomes were fixed due to the high selection pressure. Within these genomic regions were identified various genes associated with reproduction (*SLC9C1*, *PTPN12*), milk production (*IGF1*, *ABCG2*), beef production (*IFRD1*, *PTPN4*), developmental processes (*FMNL2*, *GLI2*), immune system (*CD96*, *CSK*) and coat colour (*KIT*). These selection signals detected in the genome of Slovak Spotted cattle confirm the constant effort of breeders to preserve the dual-purpose nature of this breed.

Keywords: autozygosity; effective population size; effect of selection; inbreeding; local cattle breed

Supported by the Slovak Research and Development Agency (Project No. APVV-14-0054 and No. APVV-17-0060) and MŠMT ČR (No. LTAUSA19117) the Operational Programme Integrated Infrastructure (within the Project No. 313011W112: Sustainable smart farming systems taking into account the future challenges), cofinanced by the European Regional Development Fund.

https://doi.org/10.17221/265/2020-CJAS

The autochthonous Slovak Spotted cattle is officially recognised in Slovakia since 1958. Several native autochthonous breeds (grey-brown Carpathian cattle, red Carpathian cattle and grey Steppe cattle) contributed to its breeding history, mainly via crossbreeding in the 19th century. After 1972, the main breeders' request was to increase milk production by grading-up using dairy breeds such as Jersey, Ayrshire and Holstein. The Slovak Spotted cattle belong to the dual-purpose group of breeds, with high milk and beef production. Economic efficiency of the breed is based on the ability to digest high fibre diets related to high production, regular fertility, earliness of maturity, longevity and adaptation. Special interest is given to the solid, correct type of udder, legs, claws and good musculature (Association of Slovak Spotted Cattle Breeders – Cooperative 2020).

Dual-purpose breeds, including Slovak Spotted cattle, had to adapt to the production environment; therefore, it was challenging to maintain a balance between the milk and beef production traits and reproduction in breeding programs. Generally, breeding strategies for dual-purpose breeds were focused in the past on efficient production accompanied by a rapid increase of genetic gain to achieve breeding objectives and raise profitability. These gains have been achieved by intense selection of sires and dams combined with broad use of artificial insemination and animal recording at the level of the entire population. This could reduce genetic diversity, mainly due to the mating of relatives (Amaya et al. 2020).

Genetic diversity represents the variety of genotypes and alleles in a population reflected in morphology, physiology, and behaviour differences between individuals within a population (Frankham et al. 2002). Preservation of genetic diversity is essential, mainly due to its significant contribution to animal production sustainability. The loss of genetic diversity could lead to a decrease of viability within a population (Makanjoula et al. 2020). The genetic diversity in Slovak Spotted cattle was described by several authors at the level of pedigree and molecular data (Hazuchova et al. 2013).

In population genetics, various sophisticated tools can be applied to describe the state of genetic diversity within and across populations. The effective population size (N_e) and inbreeding coefficient (F) belong to the most important population genetics parameters. It is generally accepted that mating of close relatives leads to a reduction in genetic diversity and fitness due to inbreeding depression. The inbreeding level most often increases because of the use of a limited number of sires and artificial insemination over past generations (Zhang et al. 2015). Studies using dense molecular markers throughout the genome are currently conducted to calculate more accurate estimates of inbreeding intensity and provide information about realised inbreeding in a particular population. Genomic inbreeding (F_{ROH}) can be estimated based on the distribution of homozygous segments in the genome, so-called runs of homozygosity (ROH). Runs of homozygosity are defined as contiguous homozygous regions in the genome where the two haplotypes inherited from the ancestors are identical by descent (IBD) (Bjelland et al. 2013). Calculation of F based on the genome-wide ROH distribution is a more informative approach than genealogic or other estimates because it firmly correlates with the overall homozygosity of autosomal genome derived from common ancestors (McQuillan et al. 2008). Ferencakovic et al. (2013) reported that the ROH segments of 2–4 Mb represented the distant generations of the common ancestors (25~13 generations ago), which cannot usually be captured using pedigree information. The ROH segments longer than 8 Mb represent the proportion of autozygosity originating from ancestors born 6–7 generations ago. The ROH longer than 16 Mb reflect ancestors that were born 3–6 generations ago (Moravcikova et al. 2017). The ROH patterns in the cattle genome are affected mainly by the artificial selection of superior animals. The artificial selection increases the proportion of homozygous genotypes close to the target loci involved in the genetic control of desired phenotypic traits (Szmatola et al. 2016). Thus, changes in certain regions containing functional genetic variants are the outcomes of the intense selection pressure, in a form of so-called selection signals. The identification and definition of such selection signals resulting from the ROH distribution can provide not only a basic understanding of evolutionary changes shaping the genome, but also can be very useful for identifying domestication related loci that may help to support the genetic improvement of economically important traits (Szmatola et al. 2016).

Effective population size could be explained in more directions. According to Wright (1938), the N_e was defined as the size of the idealised popu-

https://doi.org/10.17221/265/2020-CJAS

lation that would show the same level of inbreeding or allele frequency distribution under genetic drift as the evaluated population. Thus, the classical evolutionary theory of effective population size is based on the fluctuation of the gene frequency variance or increase in inbreeding (Wright 1938). Different methods were designed to evaluate the effective population size, depending on the pedigree and molecular data. If genomic data are available for a population, its effective size can be derived from the random linkage disequilibrium (LD) that arises by chance in each generation in a finite population. Barbato et al. (2015) developed a method to evaluate the N_e at different time periods using LD from several densely spaced markers on the chromosome segments. The effective population size derived from linkage disequilibrium (N_{eLD}) is estimated based on the strength of LD between loci with different genetic distances in the genome. In addition, the extent of LD in the genome can be used to explain the effect of breeding strategies on the frequency of specific alleles controlling desirable phenotypic traits and can provide insight into crossbreeding levels in populations. The LD level is influenced mainly by physical proximity and many other factors that lead to a change in the frequency of alleles, including genetic drift, mutation, migration, natural and artificial selection (Jasielczuk et al. 2016).

The aim of this study was to estimate effective population size, calculate the level of genomic inbreeding and analyse the ROH segment distribution in the genome of Slovak Spotted cattle. Subsequently, the effect of artificial selection on genome architecture was determined by identifying selection signals and the description of biologically important protein-coding genes located in the genomic regions with the highest signals.

MATERIAL AND METHODS

Data and quality control

The analysed data set of Slovak Spotted cattle consisted of 37 sires and 48 dams genotyped using two different DNA microarrays, Illumina BovineSNP50 BeadChip (Illumina, Inc., San Diego, CA, USA) in case of sires and ICBF International Dairy and Beef (IDB, Mullen et al. 2013) in case of dams, in commercial laboratories (GeneControl GmbH, Poing, Germany; Weatherbys Ireland GSB Ltd., Naas, Ireland). Analysed animals were selected to represent the nucleus of the population in Slovakia (AI bulls and bull mothers are ancestors of overall 5 968 animals in the active population of the breed). Due to different genotyping platforms, the consensus map including only autosomal single nucleotide polymorphism (SNP) markers common to both datasets was first created. Then, all individuals and SNP markers with a proportion of missing data higher than 10% were removed. Only polymorphic markers with minor allele frequency (MAF) higher than 1% were chosen for subsequent analyses. Data cleaning was performed by PLINK v1.9 (Chang et al. 2015).

Estimation of inbreeding coefficient

The ROH segments were defined according to Ferencakovic et al. (2013) as genomic regions with 36 or more consecutive homozygous calls with the maximum gap between consecutive SNPs of 1 Mb and minimum density of one SNP per every 100 kb. The minimum number of SNP markers in ROH segments (l) was calculated according to Lencz et al. (2007) to reduce the number of false-positive ROH as follows:

$$l = \frac{\log_e \alpha / n_s \times n_i}{\log_e (1 - \overline{het})} \quad (1)$$

where:

n_s – the total number of SNP markers in the dataset;
n_i – the number of individuals;
$\overline{het}$ – the average SNP heterozygosity;
α – the percentage of false-positive ROH segments; α was set to 0.05.

Based on the concept proposed by Fisher et al. (1954), there is a theoretical relationship between the length of IBD fragments and the number of generations from a common ancestor. Therefore, the ROH distribution within the genome was evaluated separately for the five categories depending on their length (ROH > 1 Mb, > 2 Mb, > 4 Mb, > 8 Mb and > 16 Mb). The number of missing calls allowed depended on the ROH length: 0 for ROH > 1 Mb and > 2 Mb, one missing call for ROH > 4 Mb, two missing calls for ROH > 8 Mb and four missing calls for ROH > 16 Mb. One heterozygous call was allowed only for the ROH length > 16 Mb. The total number of ROH, the average length of ROH (in Mb)

https://doi.org/10.17221/265/2020-CJAS

and the sum of all ROH segments per animal was calculated for each length category.

The genomic measure of individual autozygosity or genomic inbreeding (F_{ROH}) was defined as a proportion of autosomal genome located in ROH of specific minimal length relative to the overall genome. The genomic inbreeding was calculated using the following formula (McQuillan et al. 2008):

$$F_{ROH} = \frac{\Sigma L_{ROH}}{L_{autosome}} \quad (2)$$

where:

ΣL_{ROH} – the length of all ROH in the genome of an individual, which simultaneously contains a minimum specified number of consecutive homozygous SNPs;

$L_{autosome}$ – the specific length of the autosomal genome covered by the markers on the SNP microarray.

Effective population size based on linkage disequlibrium

The historical and current N_e was estimated by the SNeP software (Barbato et al. 2015), which works on the principle of relationships between LD, N_e and the recombination rate. The historical effective population size was expressed as a function of time and genetic distance between loci assuming the constant linear growth of N_e with the time expressed by past generations.

Historical N_e was estimated for 50 generations ago, which correspond to the breeding history of Slovak spotted cattle in Slovakia. The high LD value within interconnected SNPs reflects the population development in the past, while the low LD value among distant SNPs points to recent history (Kukuckova et al. 2017). The N_e values were estimated as follows (Barbato et al. 2015):

$$N_e = \frac{1}{4f(c_t)}\left(\frac{1}{E[r^2_{adj}|c_t]} - \alpha\right) \quad (3)$$

where:

N_e – the effective population size;

c_t – the recombination distance, whereas t refers to the time in past generations;

r^2_{adj} – the adjusted estimation of LD based on the sample size (proportional to the physical distance between the SNPs);

α – the mutation rate adjustment; α was set to 2.2.

The current N_e was derived based on the linear regression performed on estimates obtained for the past 50 generations (N_{eLD} from 10 to 50), according to Kukuckova et al. (2017).

Selection signatures resulting from ROH distribution

Analysis of genome-wide selection signals was based on a presumption that the identified autozygosity islands across the analysed breed genome are a result of selective breeding for traits of interests. The selection signals were characterised by SNPs with extreme frequency in ROH segments longer than 4 Mb and determined based on the calculation of runs incidence per each SNP using PLINK v1.9 (Chang et al. 2015). The visualisation of selection signals was based on the frequency (%) of overlapping ROH shared among samples. The cut-off value reflecting regions significantly affected by intense selection pressure was set based on the upper quartile of a boxplot.

The protein-coding genes located in genomic regions showing selection signals were determined using the Ensembl database (www.ensembl.org; Ensembl release 103 – February 2021). The biological function of identified genes was analysed by Over-Representation Analysis (ORA) implemented in the web-based toolkit WebGestalt (www.webgestalt.org). For this analysis, the functional databases of gene ontology (GO) and biological processes were selected. The criteria for statistically significant genes involved in the biological processes were as follows: genome protein-coding reference set was chosen by setting a minimum of 20 genes for a GO category, all categories less than the selected number were removed; significance level was 10, that means the GO categories were first ranked based on the false discovery rate 0.05 and then the top 10 most significantly enriched categories were selected.

RESULTS AND DISCUSSION

After applying the quality control, the dataset consisted of 85 animals and 37 833 SNPs that covered 2 644 825.88 kb of the autosomal genome. The average distance between adjacent autosomal markers was 69.40 kb. Two animals and 933 SNPs were

https://doi.org/10.17221/265/2020-CJAS

filtered out due to a low call rate, and 1 294 SNPs did not meet the MAF limit value.

Across the autosomal genome of Slovak Spotted cattle, the average MAF was 0.26 ± 0.14. Slovak Spotted cattle are bred as a population open to the import of genetic material from other European Simmental populations; therefore, the observed average MAF is comparable with other populations (Ilie et al. 2020).

The average observed heterozygosity at a level of 0.35 ± 0.15 was comparable with previous studies of Jasielczuk et al. (2016) and Amaya et al. (2020) in Simmental cattle. Heterozygosity may vary depending on different effects according to the farmers' mating decisions and mainly a mixture of imported sire genetics as a significant factor affecting diversity. The highest proportion of heterozygous genotypes was found on chromosome 21, where nine SNPs showed 25% heterozygosity.

Estimation of inbreeding coefficient

The inbreeding coefficient derived from ROH in different length categories was used to differentiate between ancient and recent inbreeding. The highest proportion of the genome covered by ROH and the average total length were found for category ROH > 1 Mb (Table 1). The lowest proportion of the genome covered by ROH segments was recorded for category ROH > 16 MB (0.41%) that explain recent inbreeding derived from the common ancestors born 3–6 generations ago. As presented in the study based on pedigree data, AI sires show a higher level of inbreeding (0.76%) compared to the entire population of Slovak Spotted cattle (0.36%) (Hazuchova et al. 2013). Thus, the comparison of the pedigree and genomic-based estimates of inbreeding in the current generation confirmed that animals selected for the analysis are a representative sample of the analysed population. Generally, pedigree inbreeding of Slovak Spotted cattle was decreasing in the inbred population, whereas increasing in the entire population due to the presence of common ancestors in past generations. Higher inbreeding in the population (1991–2000) could be an artefact of the use of older sires, being more frequent in the pedigrees. Thus, obtained results reflected the long-term effort of breeders to reduce inbreeding in the population of Slovak Spotted cattle. The summary statistics of F_{ROH} within each analysed category are listed in Table 1.

Table 1. Summary of results for each category of runs of homozygosity (ROH) segments

ROH category	Average No. of ROH ± SD	Average length (Mb) ± SD	F_{ROH} (%)
ROH > 1 Mb	16.75 ± 7.23	72.28 ± 43.64	2.73 ± 1.65
ROH > 2 Mb	15.39 ± 6.80	69.87 ± 43.19	2.64 ± 1.63
ROH > 4 Mb	6.44 ± 3.63	47.39 ± 37.53	1.79 ± 1.41
ROH > 8 Mb	1.62 ± 1.51	22.29 ± 29.83	0.84 ± 1.12
ROH > 16 Mb	0.44 ± 0.89	10.75 ± 27.63	0.41 ± 1.04

F_{ROH} = genomic inbreeding for each category of ROH

In the last few years, various studies analysing F_{ROH} have been realised in dual-purpose cattle. Moravcikova et al. (2017) reported for Slovak Pinzgau cattle genomic inbreeding at a level of 0.81%, which is a higher value compared to Slovak Spotted cattle. Signer-Hasler et al. (2017) found in the population of 248 Swiss Simmental sires a higher average sum of the lengths of ROH segments (96.6 Mb) and a lower level of inbreeding in the current population (F_{ROH} = 0.039) compared to our results.

Effective population size based on linkage disequilibrium

The extent of linkage disequilibrium was estimated separately for each autosomal chromosome. The average squared coefficient of correlation (r^2_{LD}) values expressing LD for each autosome are listed in Table 2. The highest average LD represented by r^2_{LD} was found for chromosomes 2 and 5, while the lowest was shown by chromosomes 18 and 27. The highest average r^2_{LD} for chromosome 5 was previously reported in Simmental cattle also by Jasielczuk et al. (2016). It has been shown that the different level of LD in particular genomic regions is related to the effect of selective breeding on the frequency of specific alleles involved in the genetic control of desired phenotypic traits.

Figure 1 shows the N_e estimates for Slovak Spotted cattle across 50 generations. The estimates of the historical effective population size showed a linear decrease, which could be a result of the loss of genetic variability through generations. The obtained linear regression function indicated that the current effective population size is 57.96 animals and the decrease in the number of animals per generation is at the level of 6.88 (r^2 = 0.99). Pedigree-based N_e estimates of sires (252.93) and dams

https://doi.org/10.17221/265/2020-CJAS

Table 2. Summary statistics for the level of linkage disequilibrium across the autosomal genome and average physical distance between single nucleotide polymorphisms (SNPs)

Chromosome	Average r^2_{LD}	SD	Average distance between SNPs (Mb)	SD
1	0.050	0.061	4.645	2.931
2	0.052	0.064	4.612	2.923
3	0.047	0.055	4.670	2.930
4	0.049	0.061	4.629	2.928
5	0.052	0.062	4.599	2.903
6	0.050	0.064	4.594	2.912
7	0.047	0.061	4.627	2.943
8	0.045	0.054	4.649	2.919
9	0.048	0.059	4.598	2.915
10	0.047	0.055	4.587	2.912
11	0.047	0.058	4.543	2.919
12	0.048	0.057	4.430	2.906
13	0.045	0.053	4.603	2.925
14	0.047	0.056	4.502	2.891
15	0.047	0.054	4.671	2.933
16	0.049	0.061	4.566	2.914
17	0.049	0.058	4.511	2.906
18	0.044	0.053	4.489	2.919
19	0.045	0.053	4.620	2.919
20	0.046	0.053	4.461	2.879
21	0.050	0.061	4.574	2.893
22	0.046	0.056	4.492	2.899
23	0.048	0.057	4.507	2.910
24	0.046	0.054	4.566	2.906
25	0.046	0.051	4.511	2.881
26	0.047	0.055	4.595	2.897
27	0.044	0.050	4.509	2.919
28	0.045	0.051	4.534	2.936
29	0.048	0.054	4.457	2.893

r^2_{LD} = average squared coefficient of correlation expressing LD

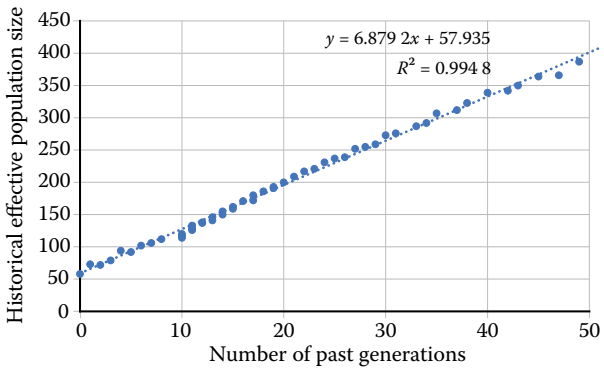

Figure 1. The estimation of effective population size trends across generations based on the linear regression

(528.88) (Hazuchova et al. 2013) overestimated the state of the population diversity. One of the reasons could be a limited number of generations traced in the pedigrees and pedigree completeness or the fact that presented results reflect another 20 years of population development. The Slovak Spotted cattle is managed as a cosmopolitan breed, thus a lower number of local AI sires is present in the mating program, which could negatively influence the gene pool of the breed in the long term. The population is open to introduction of dual-purpose Simmental breeds from all over the world. The rapid decrease in the genome-based effective population size in the current generation can be explained mainly by the change of farmers' preferences, use of a limited number of highly selected sires, changes of a production system, thus increasing the risk of diversity loss.

Selection signatures based on ROH segment distribution

Analysis of the genome-wide data of Slovak Spotted cattle showed 579 homozygous SNPs for the ROH length category > 4 Mb. This length category corresponds to autozygosity derived from common ancestors born approximately 12.5 generations ago. These segments (> 4 Mb) reflect the history of grading up of the population, whereas shorter segments (> 1 Mb, > 2 Mb) are connected with the phylogeny of the breed. The upper quartile set the threshold for the selection signals to 4%. Across the autosomal genome, 21 selection signals located within 12 different chromosomes were identified (Table 3, Figure 2). The homozygous regions had an overall length of 106.05 Mb. The lowest number (18) of homozygous SNPs was observed on chromosome 2, while the highest proportion was found on chromosome 6 (218). The longest ROH segment was observed on chromosome 6 with a start position of 29.08 Mb and an end position of 43.21 Mb. The second-largest segment with an extreme frequency of SNPs in ROH was found on chromosome 5 (60.37–70.18 Mb). The shortest ROH segment was found on chromosome 2 with start position 35.90 Mb and end position 37.36 Mb. A detailed description of identified selection signals represented by the genomic regions under strong selection pressure is listed in Table 3.

Compared to this study, Szmatola et al. (2016) observed for Simmental cattle overlapping autozy-

https://doi.org/10.17221/265/2020-CJAS

Table 3. Description of genomic regions under solid selection pressure

BTA	Start position (Mb)	End position (Mb)	Length (Mb)	No. of SNPs	No. of genes
1	55.78	61.20	5.42	75	49
	102.57	105.75	3.18	38	11
2	35.90	37.35	1.45	18	13
	42.90	44.76	1.86	21	18
	55.85	58.58	2.73	48	8
	71.13	74.13	3.00	42	36
4	41.18	48.82	7.64	124	66
5	60.37	70.18	9.81	123	100
6	21.47	23.78	2.31	39	28
	29.08	43.21	14.13	218	61
	68.12	72.97	4.85	66	55
7	40.99	46.03	5.04	59	211
9	1.33	6.57	5.24	87	3
	49.81	53.26	3.45	65	27
11	91.70	97.09	5.39	68	109
	99.04	100.86	1.82	32	65
12	33.78	39.02	5.24	58	56
21	19.83	24.40	4.57	84	123
	29.33	34.81	5.48	77	97
22	26.26	35.63	9.37	112	44
25	35.43	39.50	4.07	73	170

BTA = *Bos taurus* autosome; SNPs = single nucleotide polymorphisms

gosity islands on nine autosomes. They also identified strong selection signals on autosomes 2, 6, 11, and 16. Singer-Hasler et al. (2017) confirmed the strong effect of artificial selection on chromosome 6 in Swiss Simmental cattle. Cesarani et al. (2021), analysing five European populations of Simmental cattle, found the highest ROH occurrence on chromosome 6. It follows that chromosome 6 was under intense selection pressure in several Simmental populations distributed across Europe. This is probably because chromosome 6 includes the regions of interest that have been preferred by breeders for a long of time, and therefore they are inherited from generation to generation.

A total of 663 candidate genes were identified within genomic regions under selection pressure. Identified genes are involved in different molecular, biological and cellular processes associated with animal adaptation to the production environment. Selection signals were identified within genomic regions close to genes affecting reproduction (*SLC9C1*, *PTPN12*, *SRPK2*, *NAMPT*, *HBP1*, *SYCP3*, *BDH2*, *BMPR1B*, *KDR*, *SRD5A3*, *PCSK4*, *ATP8B3*, *NR5A1*, *FGF9*, *SNUPN*, *STAG3*, *SMURF1*), milk production (*IGF1*, *ABCG2*, *STAB2*, *SNCA*, *HERC6*, *KCNIP4*, *HERC5*, *MFGE8*), beef production (*IFRD1*, *PTPN4*, *RIC8B*, *PDGFRA*, *CRTC3*, *PLOD3*), developmental processes (*FMNL2*, *GLI2*, *PPP3CA*, *GPRIN3*, *SPP1*, *ADAMTSL3*, *FOXP1*), immune system (*CD96*, *LY75*, *NFKB1*, *RNF126*, *CSK*) and coat colour (*KIT*).

The strongest selection signals were observed on chromosome 5 and chromosome 6. On chromosome 5 are located two important protein-coding genes; the *SYCP3* gene associated with spermato-

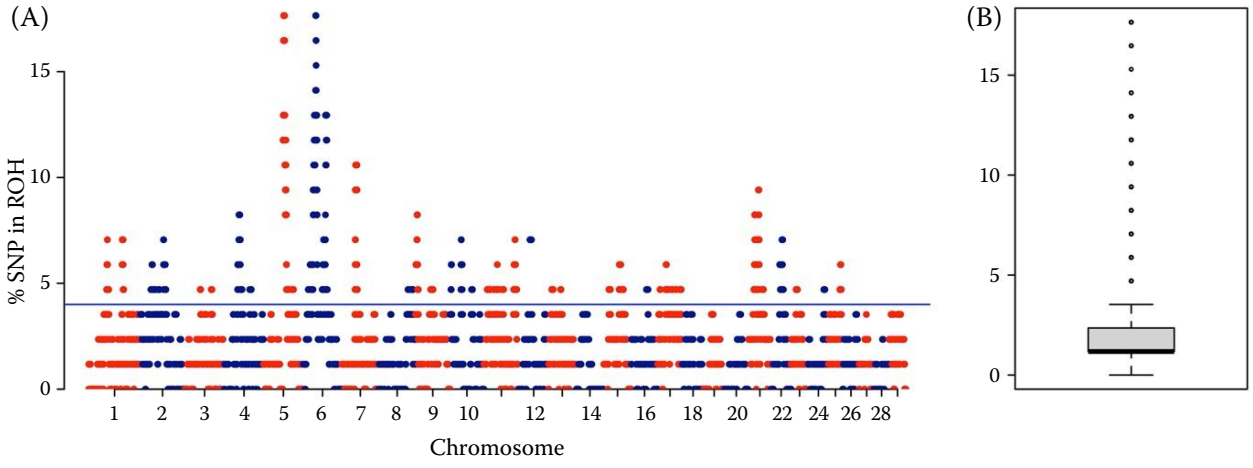

Figure 2. (A) Frequency of single nucleotide polymorphisms (SNPs) in runs of homozygosity (ROH) > 4 Mb in the genome of Slovak Spotted cattle; blue line represents the threshold of significant selection signatures, dots represent the frequency of particular SNP in ROH. (B) Box plot distribution of frequency of SNPs in ROH

https://doi.org/10.17221/265/2020-CJAS

genesis (Zhou et al. 2018) and the *IGF1* gene affecting milk fatty acid composition in dairy cattle (Li et al. 2016). On *Bos taurus* autosome (BTA) 6 were identified various genes associated with several phenotypic traits; the *KIT* gene responsible for coat colour (Fontanesi et al. 2010), the *BMPR1B* gene essential for ovulation and follicular growth (Marchitelli and Nardone 2015), the *SPP1* gene associated with muscle development and carcass weight (Matsumoto et al. 2019), the *KDR* gene important for reproduction (Randhawa et al. 2016) and the *SRD5A3* gene affecting the development of the male reproductive system (Bahbahani et al. 2018).

Several significantly enriched GO terms were identified for genes located inside regions affected by selection in the Slovak Spotted genome. They have been linked with various biological processes such as smooth muscle tissue development (GO: 0048745), ovulation cycle process (GO:0022602) or cell development (GO:0048468). The list of all GO terms is shown in Table 4. The GO terms associated with the biological function of genes near the selection signals on chromosomes 5 and 6 are listed in Table 5. On chromosome 5, mainly genes implicated in biological pathways linked with regulation of hydrolase activity, transport of a solute across the lipid bilayer, actin filament organisation and spinal cord progression during development were identified. On chromosome 6, genes involved in biological pathways associated with regulation of cell adhesion, transfer of metal ions between or within cells, and the process of maintaining homeostasis were found. The protein-coding genes were identified using the Ensembl database (www.ensembl.org) and responsible biological pathways were determined by ORA implemented in the web-based toolkit WebGestalt (www.webgestalt.org).

Obtained results confirm those of Kim et al. (2013), who demonstrated that the analysis of ROH patterns could provide an overview of the impact of selection on the frequency of specific genetic variants in different genomic regions, thus, it may reply how the direction of selection shaped the genome in the past. Kim et al. (2013) reported that the association between yield traits and alleles under selection is necessary due to a huge increase in milk yield resulting from breeding programs used in commercial cattle productions. The breeding objectives of dual-purpose breeds are generally similar, and the difference may be only in production environment and population management. This fact causes a slightly different intensity of selection pressures applied to specific phenotypic traits and indirectly to the selection of different genetic variants in the genome. This study showed that the selection strategy used for Slovak Spotted cattle breeding in the past led to the fixation of specific genomic variants in its genome. Thus, it is possible to assume that in the future, other characteristics related to muscle development and exterior properties will be fixed within the genome of Slovak spotted cattle. Our hypothesis that the selection signatures found in particular genomic regions arose as a consequence of the selection of common ancestors who carried preferred alleles were also confirmed by Szmatola et al. (2016).

CONCLUSION

Sustainable development and preservation of animal biodiversity force the effort of breeders to control intensity of inbreeding in the population. To protect the gene pool of the Slovak Spotted cattle as a part of national heritage in Slovakia, prevention of the loss of genetic diversity in the future is important. The scan of ROH segment distribution in the autosomal genome of Slovak Spotted cattle indicated that the level of historical inbreeding reached 2.73% ($F_{ROH>1Mb}$) whereas a recent inbreeding load was lower 0.41% ($F_{ROH>16Mb}$). This study confirmed our assumption that the SNPs with high frequency in ROH are located mainly in regions (BTA5, BTA6) associated with performance traits, body weight, calving ease and immune system. Thus, the identified selection signals in the genome of Slovak Spotted cattle reflect mainly its breeding goals focused on high productivity with the main emphasis on milk and meat production, reproduction or immune response. The low recent effective population size (N_e = 57.94) with an estimated annual decrease of 6.88 animals highlights the essential role of a precise breeding strategy to maintain a sufficient level of genetic variability in the population. The Slovak Spotted cattle are managed as a cosmopolitan breed, and thus a lower number of AI sires is present in the mating program, which could negatively influence the gene pool of the breed in the long term.

https://doi.org/10.17221/265/2020-CJAS

Table 4. Biological pathways (www.webgestalt.org) of candidate genes (www.ensembl.org) located in regions under selection pressure

Gene set	Gene ontology term	*P*-value	Genes
GO:0051580	regulation of neurotransmitter uptake	0.000 3	*DRD3, SNCA, TOR1A, ARL6IP5*
GO:0007271	synaptic transmission, cholinergic	0.002 0	*CHRNA7, CHRNA5, CHRNA3, CHRNB4*
GO:0018230	peptidyl-L-cysteine S-palmitoylation	0.002 0	*ZDHHC23, ZDHHC12, ZDHHC20, ZDHHC4*
GO:0018231	peptidyl-S-diacylglycerol-L-cysteine biosynthetic process from peptidyl-cysteine	0.002 0	*ZDHHC23, ZDHHC12, ZDHHC20, ZDHHC4*
GO:0060079	excitatory postsynaptic potential	0.002 5	*PPP3CA, SNCA, CHRNA7, CHRNA5, CHRNA3, CHRNB4*
GO:0050850	positive regulation of calcium-mediated signaling	0.002 6	*PPP3CA, PKD2, PTBP1, CIB1*
GO:0070884	regulation of calcineurin-NFAT signaling cascade	0.002 6	*PPP3CA, PTBP1, CIB1, HOMER2*
GO:0106056	regulation of calcineurin-mediated signaling	0.002 6	*PPP3CA, PTBP1, CIB1, HOMER2*
GO:0035094	response to nicotine	0.002 6	*CHRNA7, CHRNA5, CHRNA3, CHRNB4*
GO:0099565	chemical synaptic transmission, postsynaptic	0.002 8	*PPP3CA, SNCA, CHRNA7, CHRNA5, CHRNA3, CHRNB4*
GO:0045010	actin nucleation	0.003 3	*ARPC5L, ARPIN, LMOD3, ARPC1B, ARPC1A*
GO:2001023	regulation of response to drug	0.003 6	*DRD3, SNCA, GDF9, TOR1A, PDE8A, ARL6IP5*
GO:0035584	calcium-mediated signaling using intracellular calcium source	0.003 7	*KDR, HOMER2, FIS1*
GO:0001553	luteinization	0.003 7	*PDGFRA, GDF9, NR5A1*
GO:0048745	smooth muscle tissue development	0.003 7	*PKD2, STRA6, FOXP1*
GO:0072595	maintenance of protein localisation in organelle	0.003 9	*NR5A1, HSPA5, PML, KDELR2*
GO:0001504	neurotransmitter uptake	0.003 9	*DRD3, SNCA, TOR1A, ARL6IP5*
GO:0006612	protein targeting to membrane	0.004 8	*ZDHHC23, SRP72, ZDHHC12, ZDHHC20, CIB1, FIS1, ZDHHC4*
GO:0033173	calcineurin-NFAT signaling cascade	0.005 5	*PPP3CA, PTBP1, CIB1, HOMER2*
GO:0097720	calcineurin-mediated signaling	0.005 5	*PPP3CA, PTBP1, CIB1, HOMER2*
GO:0030325	adrenal gland development	0.006 5	*PDGFRA, NR5A1, STRA6*
GO:0032890	regulation of organic acid transport	0.006 5	*PLA2R1, SNCA, ARL6IP5*
GO:0022602	ovulation cycle process	0.006 5	*BMPR1B, PDGFRA, GDF9, NR5A1*
GO:0018345	protein palmitoylation	0.007 6	*ZDHHC23, ZDHHC12, ZDHHC20, ZDHHC4*
GO:2001024	negative regulation of response to drug	0.008 2	*SNCA, PDE8A, ARL6IP5*
GO:0045333	cellular respiration	0.008 4	*SNCA, NOA1A, NDUFS7, UQCRQ, NDUFA8, IDH2, IDH3A, COX5A, CYP1A2, SUCLG2*
GO:0048468	cell development	0.008 5	*BOC, DRD3, GAP43, INHBB, TFCP2L1, SLC9B2, PPP3CA, BMPR1B, ATOH1, GSX2, PDGFRA, SPINK2B, REST, BSG, EFNA2, TCF3, GDF9, UQCRQ, AFF4, CDKL3, UBE2B, UFL1, STRBP, TOR1A, NCS1, ATP8A2, IFT88, ABHD2, CIB1, VPS33B, FES, FBXO22, PEAK1, HMG20A, ISLR2, FOXP1, LMOD3, KBTBD8, PLOD3, ACHE, EPO, NYAP1, BHLHA15, RAC1, MMD2*
GO:0060078	regulation of postsynaptic membrane potential	0.009 6	*PPP3CA, SNCA, CHRNA7, CHRNA5, CHRNA3, CHRNB4*
GO:0048511	rhythmic process	0.011 0	*DRD3, BMPR1B, PDGFRA, GDF9, NR5A1, ASS1, PSPC1, SIN3A, PML, PROK2*
GO:0042493	response to drug	0.011 8	*DRD3, PPP3CA, SNCA, ABCG2, KDR, SRP72, GDF9, UBE2B, UFL1, TOR1A, ABHD2, PDE8A, HOMER2, CHRNA7, CHRNA5, CHRNA3, CHRNB4, CYP1A2, STRA6, ARL6IP5, SLC25A26*

P-value = 0.05 significance level based on the false discovery rate

https://doi.org/10.17221/265/2020-CJAS

Table 5. Descriptions of biological pathways (www.webgestalt.org) of overlapping candidate genes (www.ensembl.org) located on chromosome 5 and 6

Gene set	Gene ontology term	*P*-value	Genes
BTA5			
GO:0051336	regulation of hydrolase activity	0.007 57	*ARL1, HSP90B1, IGF1, NUAK1, RIC8B, TDG*
GO:0034220	transmembrane transport	0.030 34	*ANO4, IGF1, SLC25A3, SLC41A2, SLC5A8*
GO:0007015	actin filament organization	0.026 32	*HSP90B1, MYBC1, WASHC3*
GO:0021510	spinal cord development	0.007 29	*ASCL1, RFX4*
BTA6			
GO:0042592	homeostatic process	0.001 63	*BDH2, KDR, NNMU, PKD2, SLC39A8, SLC9B2, SNCA, SPP1, TMEM165*
GO:0045785	positive regulation of cell adhesion	0.002 42	*IBSP, KDR, MMRN1, PPP3CA*
GO:0030001	metal ion transport	0.000 26	*KCNIP4, PKD2, PPP3CA, SLC39A8, SLC9B2, SNCA, TMEM165*

BTA = *Bos taurus* autosome; *P*-value = 0.05 significance level based on the false discovery rate

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Amaya A, Burgos W, Martinez R, Ceron-Munoz M. Linkage disequilibrium, population stratification and patterns of ancestry in Simmental cattle. Arch Zootec. 2020 Mar; 69(266):148-54.

Association of Slovak Spotted Cattle Breeders – Cooperative. The history of the breed [Internet]. [Levice (Slovak Republic)]: Association of Slovak Spotted Cattle Breeders – Cooperative; 2020 Sep. [cited 2020 Dec 26]. Available from: https://www.simmental.sk/about-breed/breed-objective.html.

Bahbahani H, Salim B, Almathen F, Al Enezi F, Mwacharo JM, Hanotte O. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. PLOS One. 2018 Jan 4;13(1): 20 p.

Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 2015 Mar 20;6: 6 p.

Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J Dairy Sci. 2013 Jul 1; 96(7):4697-706.

Cesarani A, Gaspa G, Pauciullo A, Degano L, Vicario D, Macciotta NPP. Genome-wide analysis of homozygosity regions in european simmental bulls. J Anim Breed Genet. 2021 Jan;138(1):69-79.

Chang CC, Chow CC, Twllier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015 Dec 1;4(1): 16 p.

Ferencakovic M, Hamzic E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Solkner J. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J Anim Breed Genet. 2013 Aug;130(4):286-93.

Fisher RA. A fuller theory of "Junctions" in inbreeding. Heredity. 1954 Aug;8(2):187-97.

Fontanesi L, Scotti E, Russo V. Analysis of SNPs in the KIT gene of cattle with different coat colour patterns and perspectives to use these markers for breed traceability and authentication of beef and dairy products. Ital J Anim Sci. 2010 Jan 1;9(2):217-21.

Frankham R, Ballou JD, Briscoe DA. Introduction to conversation genetics. Cambridge: Cambridge University Press; 2002 Mar 14. 617 p.

Hazuchova E, Kadlecik O, Pavlik I, Kasarda R. Assessment of genetic diversity using characteristics based on the probability of gene origin in the Slovak Spotted breed. Acta Fytotech Zootech. 2013 Oct 3;16(2):31-7.

Ilie DE, Gao Y, Nicolae I, Sun D, Li J, Han B, Gavojdian D. Evaluation of single nucleotide polymorphisms identified through the use of SNP assay in Romanian and Chinese Holstein and Simmental cattle breeds. Acta Biochim Pol. 2020 Sep 17;67(3):341-6.

Jasielczuk I, Gurgul A, Szmatola T, Zabek T, Pawlina K, Semik E, Bugno-Poniewierska M. Linkage disequilibrium and haplotype block structure in Limousin, Simmental and native Polish Red cattle. Livest Sci. 2016 Sep 1;191:57-63.

Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y, Sonstegard TS. Effect of arti-

https://doi.org/10.17221/265/2020-CJAS

ficial selection on runs of homozygosity in U.S. Holstein cattle. PLOS One. 2013 Nov 14;8(11): 14 p.

Kukuckova V, Moravcikova N, Ferencakovic M, Simcic M, Meszaros G, Solkner J, Trakovicka A, Kadlecik O, Curik I, Kasarda R. Genomic characterization of Pinzgau cattle: Genetic conservation and breeding perspectives. Conserv Genet. 2017 Aug;18(4):893-910.

Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci USA. 2007 Dec 11;104(50):19942-7.

Li C, Sun D, Zhang S, Yang S, Alim MA, Zhang Q, Li Y, Liu L. Genetic effects of FASN, PPARGC1A, ABCG2 and IGF1 revealing the association with milk fatty acids in a Chinese Holstein cattle population based on a post genome-wide association study. BMC Genet. 2016 Dec;17(1): 16 p.

Makanjoula BO, Miglior F, Abdalla EA, Maltecca C, Schenkel FS, Baes CF. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J Dairy Sci. 2020 Jun 1;103(6):5183-99.

Marchitelli C, Nardone A. Mutations and sequence variants in GDF9, BMP15, and BMPR1B genes in Maremmana cattle breed with single and twin births. Rend Lincei. 2015 Nov;26(3):553-60.

Matsumoto H, Kohara R, Sugi M, Usui A, Oyama K, Mannen H, Sasazakic S. The non-synonymous mutation in bovine SPP1 gene influences carcass weight. Heliyon. 2019 Dec;5(12): 5 p.

McQuillan R, Leutenegger A, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF. Runs of homozygosity in European populations. Am J Hum Genet. 2008 Sep;83(3):359-72.

Moravcikova N, Kasarda R, Kukuckova V, Kadlecik O. Effective population size and genomic inbreeding in Slovak Pinzgau cattle. Agric Conspec Sci. 2017 Dec 18;82(2):97-100.

Mullen MP, Mcclure MC, Kearney JF, Waters SM, Weld R, Flynn P, Creevey CJ, Cromie AR, Berry DP. Development of a custom SNP chip for dairy and beef cattle breeding, parentage, and research. Interbull Bull. 2013;47:58-66.

Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW. A meta-assembly of selection signatures in cattle. PLOS One. 2016 Apr 5;11(4): 30 p.

Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, Gredler B, Bapst B, Flury C. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet Sel Evol. 2017 Dec;49(1): 13 p.

Szmatola T, Gurgul A, Ropka-Molik K, Jasielczuk I, Zabek T, Bugno-Poniewierska M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livest Sci. 2016 Jun 1;188:72-80.

Wright S. Size of population and breeding structure in relation to evolution. Science. 1938;87:430-1.

Zhang Q, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 2015 Dec;16: 11 p.

Zhou Y, Connor EE, Bickhart DM, Li C, Baldwin RL, Schroeder SG, Rosen BD, Yang L, Van Tassell CP, Liu EL. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. GigaScience. 2018 May;7(5): 13 p.

Received: October 25, 2020
Accepted: April 21, 2021
Published online: June 14, 2021