Genomic signatures of selection in cattle through variation of allele frequencies and linkage disequilibrium

Nina MORAVČÍKOVÁ (✉), Anna TRAKOVICKÁ, Ondrej KADLEČÍK, Radovan KASARDA

Department of Animal Genetics and Breeding Biology, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia

✉ Corresponding author: nina.moravcikova1@gmail.com

ABSTRACT

The aim of this study was to evaluate the impact of artificial selection on Slovak Spotted and Slovak Pinzgau genomes through identification of selection signatures and to characterize most important genomic regions reflecting the selective breeding for traits of interest during the formation of those breeds. The genotyping data for in total of 236 animals were included in this study. Two approaches were used to identify genomic footprints of selection: Wright's F_{ST} statistic and variation in genome-wide linkage disequilibrium patterns between selected populations. Based on applied methods, in total of 18 genomic regions under strong selection pressure were detected across 10 autosomes (BTA 4, MTA5, BTA6, BTA7, BTA11, BTA12, BTA20, BTA22, and BTA23). The longest region was identified on BTA6 close to genes affecting milk production and coat colour pattern, while the shortest one was found on BTA11. In addition, inside the identified regions some of the other genes affecting the milk production traits (casein family, HAL, IGF1, ABCG2, SPP1), carcass traits and body composition (MYBPC1, MYH9, PACRGL), reproduction (AMDHD1), temperament (SNRPF), and coat colour (KIT, KDR) were found. Because of this, all of the detected regions can be attributed mostly to improvement of milk production and muscle development, thus selection for dual-purpose performance.

Keywords: dual-purpose cattle, high-throughput data, linkage disequilibrium, selection signatures, Wright's statistics

Introduction

The identification of genomic regions affected by natural and artificial selection is high of relevance in livestock population genetics. Identifying the selection signatures gives an insight into the history of selection for economically important traits and genetic adaptation to specific environments of the populations under consideration (Ma et al., 2015; Taye et al., 2017). Moreover, the elucidation of these signatures of selection may help to further genetically improve economically important breeds (Qanbari and Simianer, 2014).

The availability of huge amount of genomic data resulting from high-throughput genotyping and NGS sequencing as well as development of improved statistical tools allows to apply various highly sophisticated analyses for identifying the footprints of selection. Generally, selection signatures can be detected through the variation of allele frequencies in a certain genomic regions and decay of linkage disequilibrium (LD). When a part of the genome that confers enhanced fitness or productive ability is preferentially kept in a population by increasing the frequency of favourable alleles, neutral loci that surround this region and that are in LD with it, are also retained, thus driving the frequency of particular haplotypes in the region towards fixation in a pattern that decays progressively with distance from the causative location. Such signals of selection can be detected for example based on the reduced haplotype diversity and different LD pattern compared

Short communication

DOI: /10.5513/JCEA01/20.2.2552

Moravčíková et al.: Genomic signatures of selection in cattle through variation of allele...

to those of the surrounding background (Pérez O'Brien et al., 2014). This approach is included for example in extended haplotype homozygosity (EHH) test (Sabeti et al., 2002) and integrated haplotype score (iHS) statistics (Voight et al., 2006). On the other hand, signatures of selection can be alternatively quantified also by the FST statistic (Weir and Cockerham, 1984) as locus-specific allele frequencies resulting from genetic differentiation between populations. The F_{ST} index provides information on the genomic variation at a locus among populations relative to that within populations. Thus, F_{ST} represents also a test for evidence of selection i. e. high F_{ST} values indicate local positive adaptation while low F_{ST} values suggest negative or neutral selection (Zhao et al., 2015).

The aim of this study was to analyse the impact of artificial selection on Slovak Spotted and Slovak Pinzgau genomes through identification of selection signatures and to characterize genomic regions reflecting the selective breeding for traits of interest during the formation of those breeds.

Materials and methods

The genotyping data for in total of 236 animals, representing the nucleus of both Slovak Spotted and Slovak Pinzgau cattle, were included in this study. The sample of Slovak Spotted cattle consisted of 37 AI sires and 48 dams that were genotyped using two platforms, Illumina BovineSNP50v2 BeadChip and ICBF International Dairy and Beef v3, respectively. In case of Slovak Pinzgau cattle, all of animals (37 AI sires, 35 dams of sires, and 79 dams of dams) were genotyped by using Illumina BovineSNP50v2 BeadChip.

The data cleaning was performed by using PLINK 1.9 (Chang et al., 2015). Because of the two different genotyping platforms used for animals' genotyping, the consensus map had to be firstly constructed. The final consensus map file consisted of 40,033 markers. The quality control of genotyping data was performed to remove markers assigned to unmapped regions or with unknown chromosomal position and SNPs positioned to sex chromosomes. In the subsequent SNP pruning only samples with lower than 10% of missing genotypes, autosomal SNPs with call rate higher than 90% and minor allele frequency higher than 1% were retained in the dataset.

Two approaches were used to identify selection signatures in the genomes of Slovak Spotted and Slovak Pinzgau cattle: Wright's F_{ST} statistic and variation in genome-wide linkage disequilibrium patterns between populations. The F_{ST} values were calculated according to Weir and Cockerham (1984) for each syntenic loci by using PLINK 1.9 (Chang et al., 2015) and then averaged over 10 consecutive SNPs using sliding windows approach. The genome-wide significance threshold for regions under selection were determined based on the corresponding boxplot distribution. The differences in genome-wide LD patterns between populations were quantified by using VarLD software (Ong and Teo, 2010) over sliding windows of 50 SNPs. Genomic regions under selection were defined by the top 0.01, 0.1, 1, and 5 percentiles of signals. For identifying genes located directly in the identified regions the Genome data viewer of the bovine genome UMD3.1.1 was used (https://www.ncbi.nlm.nih.gov/genome/gdv/).
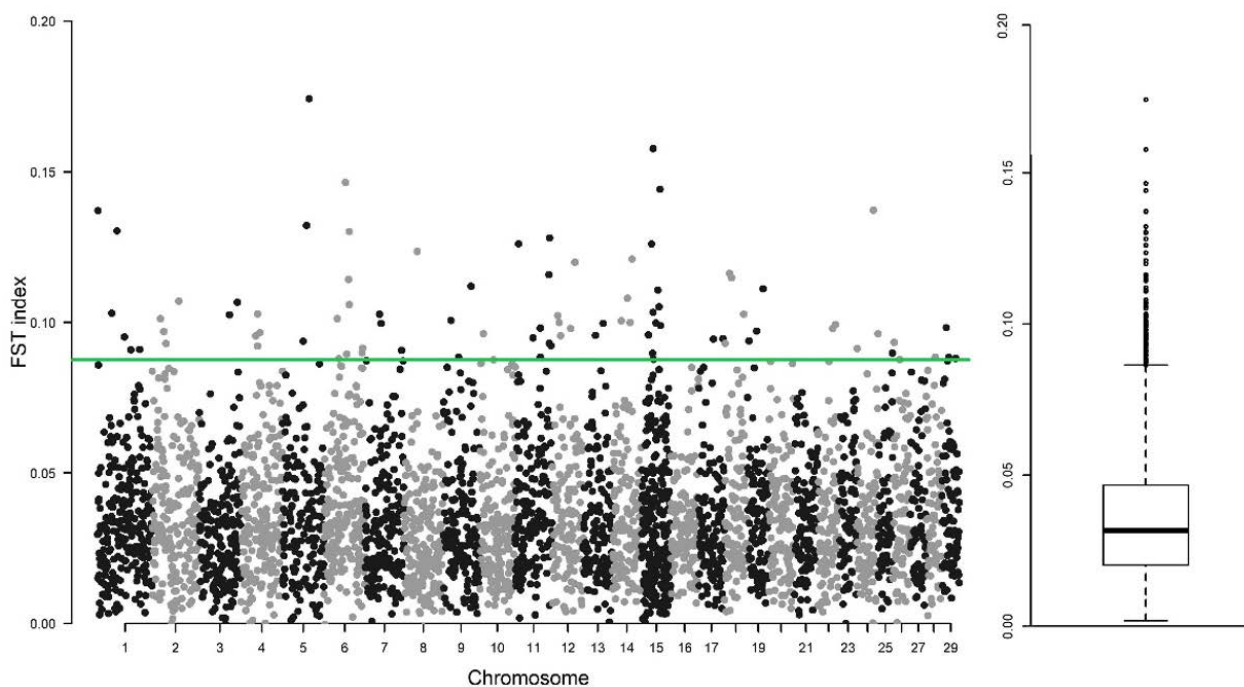
Results and discussion

After SNP pruning, overall 39,261 loci covering 2.49 Gb of the Slovak Spotted and Slovak Pinzgau genomes were retained in dataset. The average distance between adjacent syntenic SNPs was 63.67 kb. The minimum distance between SNPs was 0.02 kb, while the maximum distance was 4,428.95 kb. Based on applied approaches, in total of 18 genomic regions under strong selection were detected across 10 autosomes (Table 1).

According to the boxplot distribution the genome-wide significance threshold for genomic regions derived from Wright's F_{ST} statistics was set to 0.09 (Figure 1). The strongest signals of selection were found within 8 genomic regions distributed across three autosomes (BTA5, BTA6, and BTA15). The longest region was identified on BTA6 (62,085,065-73,092,782 bp), while the shortest one was found on BTA5 (59,598,727-60,642,347 bp). Various

Short communication

DOI: /10.5513/JCEA01/20.2.2552

Moravčíková et al.: Genomic signatures of selection in cattle through variation of allele...

Table 1. Genomic regions showing the strongest signal of selections across applied approaches

BTA	Start position (Mb)	End position (Mb)	Length (Mb)	No. of genes	QTL traits
4	44.09	46.82	2.74	27	Somatic cell score
5	59.6	60.64	1.04	36	
5	63.65	68.01	4.36	55	Dressing percentage, birth weight, longissimus muscle area, backfat EBV, follicle stimulating hormone concentration, twinning rate
5	70.31	76.92	6.61	95	
6	38.58	43.18	4.6	15	Birth weight, stature, strength, daily gain, milk yield, protein and fat percentage, protein and fat yield
6	62.09	73.09	11.01	105	Rump width, suspensory ligament, teat placement, foot angle, quality of foot and leg, quality of udder, marbling score, milk yield, protein percentage and yield, fat yield
6	109.87	111.88	2.01	11	Protein yield, fat percentage
7	42.06	44.92	2.86	110	
11	23.97	24.38	0.41	1	Yearling and weaning weight
12	25.79	27.91	2.11	10	Milk yield, protein and fat yield, ribeye muscle area, yield grade
15	30.46	35.2	4.74	51	
15	42.92	48.07	5.15	157	Estimated kidney, pelvic and heart fat
15	51.53	55.64	4.11	109	Meat tenderness
20	4.61	6.18	1.56	22	Birth weight, protein and fat yield
22	32.76	34.64	1.88	7	
23	30.44	30.97	0.53	98	Protein percentage, marbling score

Figure 1. Genome-wide distribution of F_{ST} values with corresponding boxplot

Short communication

DOI: /10.5513/JCEA01/20.2.2552

Moravčíková et al.: Genomic signatures of selection in cattle through variation of allele...

quantitative trait loci (QTLs) have been reported inside the genomic location of identified regions, including those that affect birth weight (Casas et al., 2003), stature and strength (Hiendleder et al., 2003), twinning rate (Lien et al., 2000), milk production (Velmala et al., 1999), quality of udder (Hiendleder et al., 2003), longissimus muscle area (Casas et al., 2003), marbling score (Kim et al., 2003), and meat tenderness (Keele et al., 1999).

The analysis of differences in genome-wide LD patterns between Slovak Spotted and Slovak Pinzgau populations revealed in total of 10 genomic regions distributed across autosomes BTA4, BTA5, BTA6, BTA7, BTA11, BTA12, BTA20, BTA22, and BTA23 (Figure 2).

The longest region was found on BTA5 (63,647,446-68,006,770 bp), including QTLs for longissimus muscle area (Casas et al., 2003), backfat EBV (Li et al., 2004), follicle stimulating hormone (Casas et al., 2004). The shortest region was detected on BTA11 (23,966,836-24,379,262 bp). Table 1 summarizes the subset of biologically and economically most important QTLs located within affected genomic regions. Besides them, some of the genes involved in a wide variety of biological processes, including milk production (casein family, HAL, IGF1, ABCG2, SPP1), muscle formation and body composition (MYBPC1, MYH9, PACRGL), reproduction (AMDHD1), temperament (SNRPF), and coat colour (KIT, KDR) were identified in target regions.
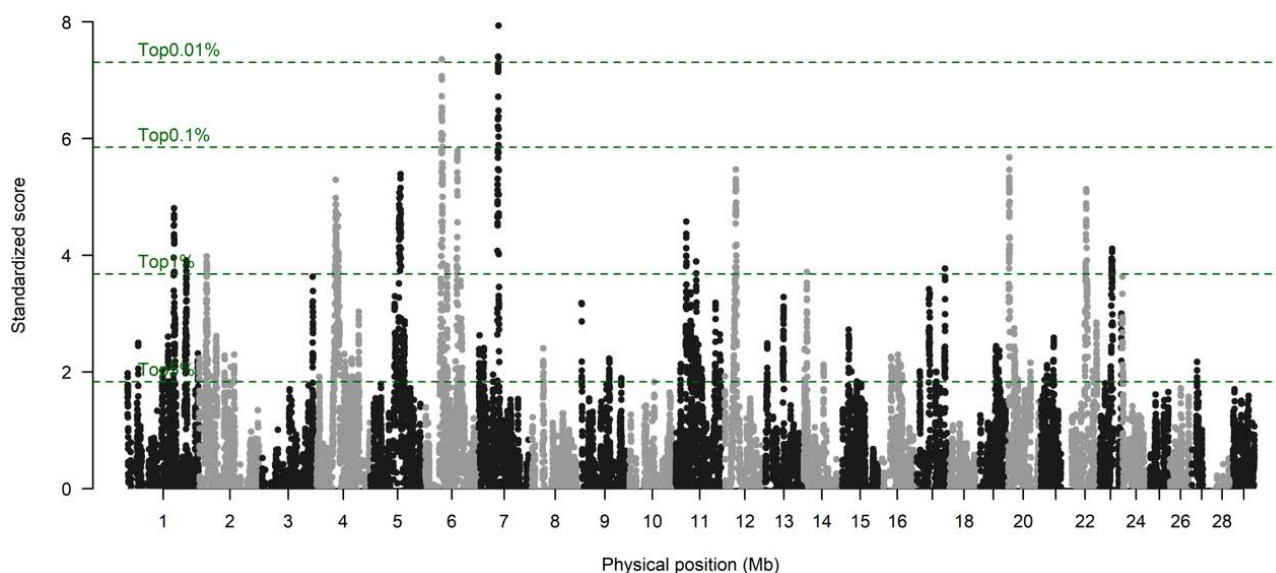

Figure 2. Differences in genome-wide LD patterns between populations (upper abline is showing the genome-wide significance threshold)

Conclusions

This study confirmed that the regions displaying selection signatures in Slovak Spotted and Slovak Pinzgau cattle are linked mostly to milk production and muscle development, thus selection for dual-purpose performance. Both of applied approaches represented efficient tools to identify the genomic regions as well as genes related to traits with biological and economical importance.

Acknowledgements

References

Casas, E., Shackelford, S.D., Keele, J.W., Koohmaraie, M., Smith, T.P., Stone, R.T. (2003) Detection of quantitative trait loci for growth and carcass composition in cattle. Journal of Animal Science, 81 (12), 2976-2983.
DOI: https://dx.doi.org/10.1046/j.1365-2052.2003.01067.x

Short communication

DOI: /10.5513/JCEA01/20.2.2552

Moravčíková et al.: Genomic signatures of selection in cattle through variation of allele...

Casas, E., Lunstra, D.D., Stone, R.T. (2004) Quantitative trait loci for male reproductive traits in beef cattle. Animal Genetics, 35 (6), 451-453. DOI: https://dx.doi.org/10.1111/j.1365-2052.2004.01190.x

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., PurcelL, S.M., Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4, 7.
DOI: https://dx.doi.org/10.1186/s13742-015-0047-8

Hiendleder, S., Thomsen, H., Reinsch, N., Bennewitz, J., Leyhe-Horn, B., Looft, C., Xu, N., Medjugorac, I., Russ, I., Kühn, C., Brockmann, G.A., Blümel, J., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Schwerin, M., Förster, M., Kalm, E., Erhardt, G. (2003) Mapping of QTL for body conformation and behavior in cattle. Journal of Heredity, 94 (6), 496-506.
DOI: https://dx.doi.org/10.1093/jhered/esg090

Keele, J.W., Shackelford, S.D., Kappes, S.M., Koohmaraie, M., Stone, R.T. (1999) A region on bovine chromosome 15 influences beef longissimus tenderness in steers. Journal of Animal Science, 77 (6), 1364-1371.
DOI: https://dx.doi.org/10.2527/1999.7761364x

Kim, J.W., Park, S.I., Yeo, J.S. (2003) Linkage mapping and QTL on chromosome 6 in Hanwoo (Korean Cattle). Asian-Australasian Journal of Animal Sciences, 16 (10), 1402-1405.
DOI: https://dx.doi.org/10.5713/ajas.2003.1402

Li, C., Basarab, J., Snelling, W.M., Benkel, B., Kneeland, J., Murdoch, B., Hansen, C., Moore, S.S. (2004) Identification and fine mapping of quantitative trait loci for backfat on bovine chromosomes 2, 5, 6, 19, 21, and 23 in a commercial line of Bos taurus. Journal of Animal Science, 82 (4), 967-972.
DOI: https://dx.doi.org/10.1093/ansci/82.4.967

Lien, S., Karlsen, A., Klemetsdal G., Våge, D.I., Olsaker, I., Klungland, H., Aasland, M., Heringstad, B., Ruane, J., Gomez-Raya, L. (2000) A primary screen of the bovine genome for quantitative trait loci affecting twinning rate. Mammalian Genome, 11 (10), 877-882.
DOI: https://dx.doi.org/10.1007/s003350010180

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., Simianer, H. (2015) Properties of different selection signature statistics and a new strategy for combining them. Heredity, 115 (5), 426-436.
DOI: https://dx.doi.org/10.1038/hdy.2015.42

Ong, R.T., Teo, Y.Y. (2010) varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics, 26 (9), 1269-1270.
DOI: https://dx.doi.org/10.1093/bioinformatics/btq125

Pérez O'Brien, A.M, Utsunomiya, Y., Mészáros, G., Bickhart, D.M., Liu, G.E., Van Tassell, C.P., Sonstegard, T.S., Da Silva, M.V.B., Garcia, J.F., Sölkner, J. (2014) Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. Genetics Selection Evolution, 46 (1), 19.
DOI: https://dx.doi.org/10.1186/1297-9686-46-19

Qanbari, S., Simianer, H. (2014) Mapping signatures of positive selection in the genome of livestock. Livestock Science, 166, 133-143.
DOI: https://dx.doi.org/10.1016/j.livsci.2014.05.003

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature, 419, 832-837.
DOI: https://dx.doi.org/10.1038/nature01140

Velmala, R.J., Vilkki, H.J., Elo, K.T., de Koning, D.J., Mäki-Tanila, A.V. (1999) A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Animal Genetics, 30 (2), 136-143.
DOI: https://dx.doi.org/10.1046/j.1365-2052.1999.00435.x

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K. (2006) A map of recent positive selection in the human genome. PLoS Biology, 4, e72.
DOI: https://dx.doi.org/10.3410/f.1023845.367919

Weir, B.S., Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358-1370.
DOI: https://dx.doi.org/10.2307/2408641

Zhao, F., McParland, S., Kearney, F., Du, L., Berry, D.P. (2015) Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genetics Selection Evolution, 47 (1), 49.
DOI: https://dx.doi.org/10.1186/s12711-015-0127-3