https://doi.org/10.17221/226/2019-CJAS

Analysis of selection signatures in the beef cattle genome

Nina Moravčíková[1], Radovan Kasarda[1], Luboš Vostrý[2,3], Zuzana Krupová[3], Emil Krupa[3], Kristína Lehocká[1], Barbora Olšanská[1], Anna Trakovická[1], Rudolf Nádaský[4], Radoslav Židek[5], Ľubomír Belej[5], Jozef Golian[5]*

[1]*Department of Animal Genetics and Breeding Biology, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic*

[2]*Department of Genetics and Breeding, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic*

[3]*Institute of Animal Science, Prague-Uhříněves, Czech Republic*

[4]*Poľnohospodárske družstvo Špačince, Špačince, Slovak Republic*

[5]*Department of Food Hygiene and Safety, Faculty of Biotechnology and Food Sciences, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic*

**Corresponding author: nina.moravcikova1@gmail.com*

Citation: Moravčíková N., Kasarda R., Vostrý L., Krupová Z., Krupa E., Lehocká K., Olšanská B., Trakovická A., Nádaský R., Židek R., Belej Ľ., Golian J. (2019): Analysis of selection signatures in the beef cattle genome. Czech J. Anim. Sci., 64, 491–503.

Abstract: This study aimed to evaluate the impact of selection on the genome structure of beef cattle through identification of selection signatures reflecting the breeding standard of each breed and to discover potential functional genetic variants to improve performance traits. Genotyping data of six beef breeds (Aberdeen Angus, Hereford, Limousin, Charolais, Piedmontese and Romagnola) were used to perform genome-wide scans for selection signatures. The approaches applied were based on an assumption that selection leads to linkage disequilibrium or to a decrease of genetic variability in genomic regions containing genotypes connected with favourable phenotypes. Thus, the selection signatures were analysed based on Wright's F_{ST} index, distribution of runs of homozygosity segments in the beef genome and determination of linkage disequilibrium variability between breeds. The number and length of detected selection signals were different depending on the breeds and methodological approaches. As expected due to the breeding goals of analysed breeds, common signals were located on autosomes 2, 6, 7, 13 and 20 close to the genes associated with coat colour (*KIT, KDR*), muscle development (*GDF9*, *GHRH*, *GHR*), double muscling (*MSTN*), meat tenderness (*CAST*) and intramuscular fat content (*SCD*). But, across the genomes of analysed breeds, unique selection signals were found as well. The subsequent analysis of those single nucleotide polymorphism markers can be beneficial for the genetic progress of studied breeds in future.

Keywords: artificial selection; local population; genomic autozygosity; linkage disequilibrium; Wright's statistic

In cattle, similarly like in other farm animal species, natural as well as artificial selection together with adaptation to different biogeographical regions and production systems influenced mainly allele frequencies of loci associated with fitness and variability of production traits of interest.

Supported by the Slovak Research and Development Agency (Projects No. APVV-17-0060 and No. APVV-17-0508), Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and Slovak Academy of Sciences (VEGA) (Project No. 1/0742/17), Ministry of Agriculture of the Czech Republic (Projects No. QK1910059 and No. MZe-RO0718-V003) and by the European Union (ITMS 26240220080).

https://doi.org/10.17221/226/2019-CJAS

Therefore selection has led to linkage disequilibrium (LD), and a decrease of genetic variability in the genomic regions localised close to favourable alleles. According to the theory of natural selection, the majority of molecular variants within and between breeds are selectively neutral, i.e. not affecting the fitness of an individual. Whenever the particular variant (newly created by mutation or already present in the genome) provides an individual with favourable fitness in comparison with others in the population, the frequency of such favourable genotype or allele will increase in this particular population, because it provides some selective advantage compared to others. This even more increases the frequency of other neutral variants localised in neighbouring genomic regions due to the hitch-hiking effect. As a result, selection signals, or areas under intense selection pressure, are produced. Such genomic regions are usually connected with a significant decrease of genetic variability and level of differentiation between populations as well as an increase of LD and interspecies divergence in the particular genomic region (Utsunomiya et al. 2015).

On the other hand, the genetic variants need not provide selective advantage against others, their frequency in the population can be increased randomly, e.g. due to genetic drift. Artificial selection has the same effect on the genome structure as natural selection when such signals appear as a result of the selection of target genetic variants, responsible for the improvement of production and reproduction traits of an individual (Utsunomiya et al. 2015). Currently, the availability of genomic and bioinformatics tools makes it possible to evaluate the effect of selection at the whole genome level. In the case of beef cattle, substantial progress was reached mainly due to Hereford genome sequencing and its comparison with more or less specialised breeds (The Bovine HapMap Consortium 2009).

The choice of a method for identifying genomic regions under selection pressure depends on the nature of the selection signals and the time during which the genome is exposed to selection. One of the most commonly used methods is Wright's F_{ST} statistics, based on the evaluation of differences in allele frequencies between populations. This method allows evaluating not only the effects of recent selection but also the selection that occurred 2000 to 3000 generations ago (Sabetti et al. 2007). Several modifications of this method using various modern technologies have already been available. Even if the comparison of those approaches and several single nucleotide polymorphism (SNP) arrays showed different estimates of F_{ST} index, this parameter is recognised as one of the most suitable for deciphering genomic signals of positive selection outgoing from genetic differentiation between observed populations.

Analysis of haplotypes provides another possible tool for identification of selection signals, e.g. by the method based on extended haplotype homozygosity (EHH) (Sabetti et al. 2007). The integrated haplotype score statistic (iHS) is another popular approach that provides a standardised measurement of EHH decrease around a certain point (e.g. SNP) from a derived allele relative to the ancestral (inherited) allele. Indicators of selection are then regions in which a slight decrease of haplotype homozygosity in case of derived allele appears (i.e. longer than expected haplotypes in case of ancestral allele) (Voight et al. 2006). For identification of selection signatures, the analysis of LD variability between and within breeds can be used as well (Salem et al. 2018). However, these selection signals have a temporary tendency in some cases because recombination may cause a change in the sequence of the selected locus faster than they are fixed in population.

In cattle genome, various selection signals reflecting domestication processes, natural or intensive artificial selection, have been reported. Most of the applied approaches were based on Wright's F_{ST} statistics, distribution of runs of homozygosity (ROH) stretches in the genome, analysis of haplotype structure and differences in allele frequencies between populations, evaluation of haplotype homozygosity via integrated haplotype score statistic (iHS), inter-population extended haplotype homozygosity (XP-EHH) and likelihood ratio test (XP-CLR) (Kukuckova et al. 2017; Moravcikova et al. 2018). As already mentioned, each of these methods has certain advantages and disadvantages. Genomic regions identified by one method could not be identified by another method even when using the same datasets. It is caused mainly by the fact that each of the methods is sensitive to a different time scale during which the selection acted in the genome. Because of this, the combination of methods is preferred to increase the reliability of the results obtained (Utsunomiya et al. 2015).

Identification of the impact of natural and artificial selection on the cattle genome structure allows

https://doi.org/10.17221/226/2019-CJAS

better understanding of the biological mechanisms through which genetic differentiation has occurred within breeds selected for different phenotypic traits and characteristics. In beef cattle, several genomic regions and genes under intense selection pressure were identified. An extensive study of The Bovine HapMap Consortium (2009) revealed that most genomic regions showing selection signals are related to milk yield, meat quality, and feed conversion. Further studies in dairy and beef cattle have consistently identified selection signals in genomic regions controlling the coat colour pattern (*MC1R* and *KIT* genes) and the size of the body frame (Randhawa et al. 2016). A strong effect of selection was proved especially in the centromere part of chromosome 14 that is involved in the genetic control of marbling and intramuscular fat content (Wiener et al. 2011). Besides, selection signals related to intramuscular fat content and feed conversion were found on chromosome 2 and 13 (Sorbolini et al. 2015). In some beef breeds, specialised breeding programs have been developed for the production of double-muscled animals. This trait is controlled by the growth differentiation factor gene (*GDF8*, also known as myostatin, *MSTN*) (Bellinge et al. 2005). A decrease in heterozygosity and an increase in LD level in the region of this gene, confirming an intense selection pressure, were found, e.g. by Wiener and Gutierrez-Gil (2009).

The objective of this study was to analyse the effect of selection on the beef cattle genome structure through identification of genomic regions under intense selection pressure and to identify new functional genetic variants to improve meat quality.

MATERIAL AND METHODS

Analysed sample and SNP pruning. The genome-wide data of overall 403 animals, available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.42tr0.2) (McTavish et al., 2013), were used for biostatistical analyses. From in total 58 dairy and beef breeds included in this database, six beef cattle breeds belonging to the *Bos taurus taurus* subspecies were selected based on the sample size: Aberdeen Angus (n = 90), Hereford (n = 98), Limousin (n = 100), Charolais (n = 53), Piedmontese (n = 29) and Romagnola (n = 29).

All animals included in the study were genotyped using the BovineSNP50v1 BeadChip (Illumina Inc., USA). Standard quality control of genotyping data was performed using PLINK 1.9 software (Chang et al. 2015) following the methodology of McTavish et al. (2013). In the first step of SNP pruning, all SNPs localised on sex chromosomes as well as markers with unknown chromosomal position according to the cattle reference genome assembly Btau 5.0.1 were excluded from the database. In the next step, all individuals and autosomal SNPs with more than 10% of missing data were filtered out. Since previous studies have shown that markers with extremely low minor allele frequency (MAF) can significantly affect the estimation of e.g. the LD level, 0.01 was chosen as the cut-off MAF value for LD analysis. This value is also in agreement with the generally accepted theory of Mendelian inheritance. After SNP pruning, the final database consisted of 43 370 markers and 399 animals.

Analysis of selection signatures in the beef cattle genome. The effect of selection on the genome structure was evaluated based on three approaches: Wright's F_{ST} index on a genome-wide level, identification of ROH distribution in the genome and determination of LD variability between breeds.

After identifying the regions showing the most reliable selection signals, these regions were characterised according to their chromosomal position (start and end position in kbp) and compared between breeds under consideration. Subsequently, the genomic regions detected were ordered by autosomes and position to consider common signals between and within breeds and to decipher regions unique for a particular breed due to the specific breeding standard. All SNPs under selection pressure were aligned to particular quantitative trait loci (QTLs) using the Bovine Genome Database (http://bovinegenome.org). A list of genes localised directly in detected regions was prepared using the Genome Data Viewer (Btau reference sequence 5.0.1) with specific emphasis on genes controlling meat quality.

Distribution of Wright's F_{ST} index in the genome. The genome-wide distribution of Wright's F_{ST} index was scanned using PLINK 1.9 (Chang et al. 2015) for all SNPs per individual chromosomes. Since negative F_{ST} values cannot be interpreted from the biological point of view, all of the SNPs showing negative values were excluded from the database.

https://doi.org/10.17221/226/2019-CJAS

The remaining F_{ST} values were then averaged for every ten consecutive SNPs. Signals of selection, in this case expressing mainly phylogenetic differences between breeds, were characterised as regions with the highest (extreme) F_{ST} values. Based on the evaluation of the variability of those values according to the quartiles and subsequent visualisation using boxplot the threshold value was set to 0.2.

Distribution of ROH segments in the genome. Analysis of ROH segment distribution reflecting the intensity of selection was based on an assumption that autozygous regions in the genome of analysed breeds are results of selective breeding, realised in the last 12 generations. ROH segments were defined as genetic regions with a minimum of 15 consecutive homozygous genotypes and a minimum density of 1 SNP per 100 kbp. Maximum distance allowed between homozygous genotypes was set to 1 Mbp. Because between the distribution of identical by descent (IBD) fragments and the number of generations derived from common ancestor exists a theoretical relationship, a minimum length of ROH segment was set to 1 Mbp. Autozygous regions were then characterised based on SNPs with an extreme frequency of ROH segments longer than 4 Mbp. These SNPs were determined by calculating their proportion in ROH using PLINK 1.9 (Chang et al. 2015), where the occurrence of SNPs in ROH was expressed as the frequency (%) of overlapping homozygous regions within the samples. The limit value at which a given marker (with an extreme frequency in ROH) was considered as selection signal was determined similarly to the F_{ST} index by assessing their variability across the quartiles.

Variability in linkage disequilibrium. The varLD (variability in linkage disequilibrium) method comparing the level of LD in particular genomic regions between two breeds was used to evaluate the variability in LD and the detection of selection signals. The Holstein cattle were used as outgroup breed (http://dx.doi.org/10.5061/dryad.42tr0.2) (McTavish et al. 2013). This method is implemented in the VarLD software (Teo et al. 2009) that is based on an assumption that due to the positive selection of some genetic variants in a particular genomic region of a specific population, the level of LD in this region will change compared to the population whose genome was not subjected to such selection. The LD values for all SNP pairs within autosomes, expressed by r^2, were calculated over sliding windows containing 50 markers. Signals of selection were identified based on the calculation of standardised varLD score (S_i'). Genomic regions significantly affected by the selection were recognised as regions containing 99.9 and/or 99.999% of signals in a given genomic region.

RESULTS AND DISCUSSION

Analysed sample and SNP pruning. After SNP pruning, the database of genotyping information included 43 370 SNPs common to six evaluated beef breeds covering 2.49 Gbp of the genome. The average distance between markers was 57.68 ± 58.66 kbp. The minimum distance between markers was 0.01 kb, while the maximum distance was 4244.10 kbp. All animals in the dataset showed a call rate higher than 90% of SNPs.

Minor allele frequency (MAF) is a commonly used parameter for the evaluation of bi-allelic marker variability. The quality control showed that 80.32% of markers in the database was polymorphic, i.e. with MAF higher than 0.01. On average, the MAF was 0.26 ± 0.13, and its distribution was not uniform in the autosomal genome (Figure 1). Average MAF differed significantly depending on the population, from 0.21 (Piedmontese) to 0.25 (Hereford).

Analysis of selection signatures in the beef genome.

Distribution of Wright's F_{ST} index in the genome. The theory of Wright's statistics shows that F_{ST} index values vary from 0 to 1, both of these extreme values expressing either total genetic similarity (F_{ST} = 0) or maximum differentiation between populations (F_{ST} = 1). Thus, selection signals in a

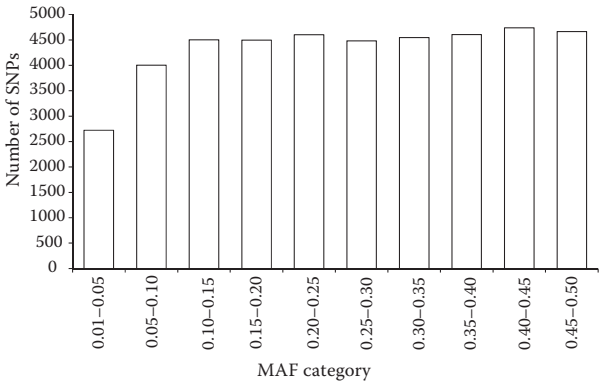

Figure 1. Distribution of minor allele frequencies (MAF) across the autosomal genome of the studied breeds

SNPs = single nucleotide polymorphisms

https://doi.org/10.17221/226/2019-CJAS

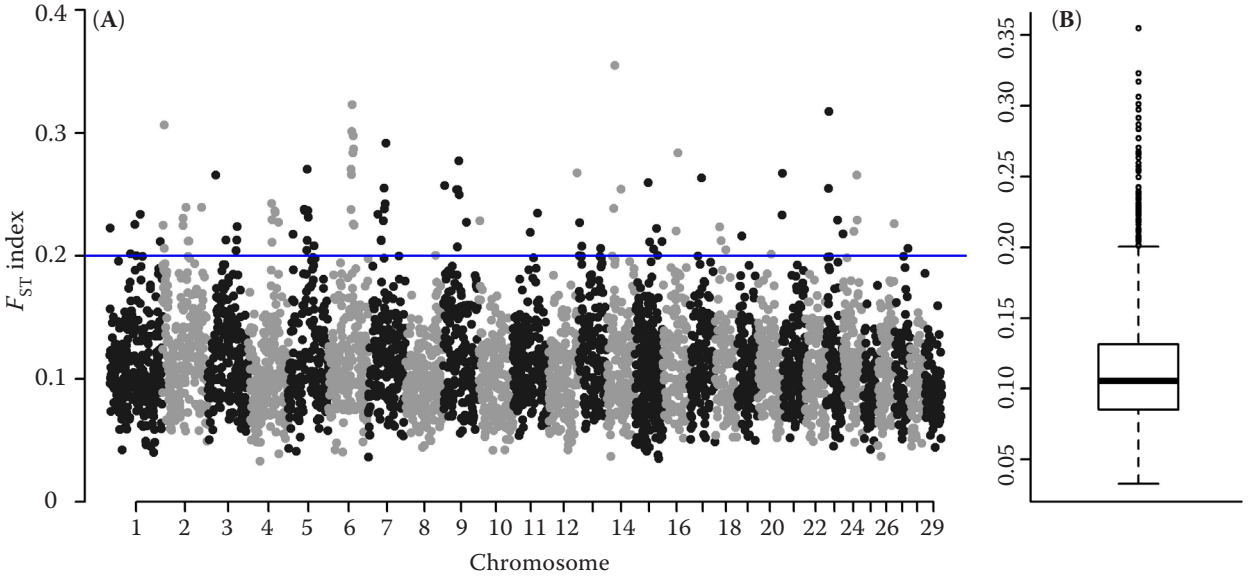

Figure 2. Distribution of F_{ST} values per autosome (**A**) and graphical distribution of F_{ST} using a box plot (**B**)

particular region in the genome can be recognised if all adjacent SNP markers show a high F_{ST} value in that region due to a different type of selection and hitch-hiking effect, or *vice versa,* if all adjacent SNPs show a low F_{ST} level due to stabilising selection between populations (Qanbari et al. 2011).

In this study, the selection signatures were recognised as autosomal regions with F_{ST} value higher than 0.2 (Figure 2). The values of F_{ST} index varied from 0.03 to 0.35, with the average of 0.11 ± 0.03. However, more than 99% of the F_{ST} values were lower than 0.25. This value is generally considered as a limit describing significant differentiation between given populations, i.e. the majority of SNP markers was under pressure of the same type of selection related to phenotypic traits typical of the beef breeds. Overall, 26 genomic regions were identified to be under significant selection pressure based on this approach. These regions were distributed across 15 chromosomes and covered 121.33 Mbp of the autosomal genome. The longest region was found on BTA9 (22.61 Mbp) and the shortest on BTA21 (0.18 Mbp). A detailed description of identified regions, including the number of genes and list of QTLs is shown in Table 1.

Within 26 selection signal regions, a set of 1604 protein-coding genes with different biological role was identified, e.g. structural, regulatory genes or genes controlling economically important traits.

In the first region on BTA1, the *MYLK* (myosin light chain kinase) gene is localised, which plays an important role in the development of muscle. Zheng et al. (2019) assigned it as a candidate gene for growth-related traits and proposed its implementation in marker-assisted selection (MAS) programmes of beef cattle. In the second region on BTA1, the *SST* gene (somatostatin gene) was identified, which is similarly like the *MYLK* gene associated with growth intensity. In the same region, the *ADIPOQ* gene (adiponectin gene) affecting marbling, the ribeye muscle area and fat thickness was also found (Zhang et al. 2014a). Another gene – *MYL6* – was allocated in the region on BTA5. This gene belongs to genes regulating muscle development in all mammals.

When looking for genes affecting cattle reproduction, in the second region on BTA1 the *SOX2* gene was identified, which affects embryonic development. On BTA7 the *GDF9* gene (bovine growth differentiating factor 9) was found, belonging to the family of growth factors that are significantly associated with the development of ovarian follicles and intensity of ovulation (Tanq et al. 2013). Previous studies showed that this family of genes is under intense selection pressure not only in beef cattle but also in dairy cattle (Bellinge et al. 2005; Wiener and Gutierrez-Gil 2009). Another gene – *FSHB*, coding beta subunit of follicle-stimulating hormone, was detected in the second region on BTA15. Several studies proved a significant effect of this gene on reproduction quality in cattle (Sharifiyazdi et al. 2018).

Similarly, like in many other studies (e.g. Randhawa et al. 2016), strong selection signals were

https://doi.org/10.17221/226/2019-CJAS

Table 1. Autosomal regions under selection pressure identified based on the distribution of Wright's F_{ST} index in the genome

BTA	Start (Mb)	End (Mb)	Length (Mb)	Genes *n*	QTL in region
1	62.046	75.434	13.388	164	live weight at age of 365 days, carcass weight, BSE resistance
	80.063	91.358	11.295	132	
2	5.277	5.664	0.387	6	
	61.736	62.266	0.530	9	marbling score, milk production, carcass weight
	70.560	77.289	6.729	45	marbling score, length of production life, birth weight, IMF
3	84.966	88.207	3.241	28	marbling score
4	70.452	71.554	1.101	10	LMA (*Longissimus thoracis* cut)
	80.303	81.434	1.131	2	
5	48.460	49.399	0.939	11	dressing percentage, birth weight, LMA, FSH concentration, EBV for backfat thickness
	55.652	56.618	0.966	38	
	57.470	60.557	3.086	154	
6	67.644	69.185	1.541	21	growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content
	70.001	72.907	2.906	35	
	74.085	76.210	2.125	6	
7	38.904	39.327	0.423	8	ovulation intensity, SCS
	45.439	48.909	3.469	70	
	51.021	53.468	2.447	67	
9	40.401	43.145	2.744	35	fat content
	46.531	69.144	22.613	148	milk yield, fat and protein yield, ham angle, conformation traits 2 and 6
14	21.560	24.573	3.014	32	fat and protein content
15	57.154	66.209	9.055	73	tenderness
16	39.553	45.018	5.464	94	carcass weight
18	14.527	18.919	4.392	60	carcass weight, dystocia (maternal), SCS
21	2.659	2.834	0.175	2	birth weight, claw angle
23	9.318	9.756	0.438	15	milk production
24	42.049	43.170	1.121	16	dressing percentage

BTA = *Bos taurus* autosome, QTL = quantitative trait locus, IMF = intramuscular fat, LMA = *Longissimus thoracis* muscle area, EBV = estimated breeding value, SCS = somatic cell score

identified close to the genes coding the coat colour of cattle on BTA6 (*KIT* and *KDR*) and BTA18 (*MC1R*). In cattle, the *KIT* gene is responsible for pied colour patterns also assigned as "spotting loci" (Fontanesi et al. 2010). The *MC1R* gene (melanocortin 1 receptor gene) determines basic coat colours (Dorshorst et al. 2015).

Distribution of ROH segments in the genome. The second approach to assessing the effect of selection on the genome structure was based on an assumption that regions with a high proportion of homozygous segments (ROH) mainly reflect the effect of selective breeding on certain phenotypic traits, defined in the breeding standards of given breeds. Resulting ROH segments located in the genome close to each other arose mainly due to the existence of alleles inherited from common ancestors that are transmitted from generation to generation. In this study, autozygous regions were defined as specific genomic regions with an extreme allele frequency in homozygous regions with a minimum length of 4 Mbp, which corresponds to the autozygosity derived from common ancestors approximately 12 generations ago. Based on the graphical visualisation of observed values (boxplot) a threshold for the identification of selection signatures was set (Figure 3). In Figure 3, it can be seen that the distribution of selection signals in the genome of the breeds of interest was not uniform and depended on the particular

https://doi.org/10.17221/226/2019-CJAS

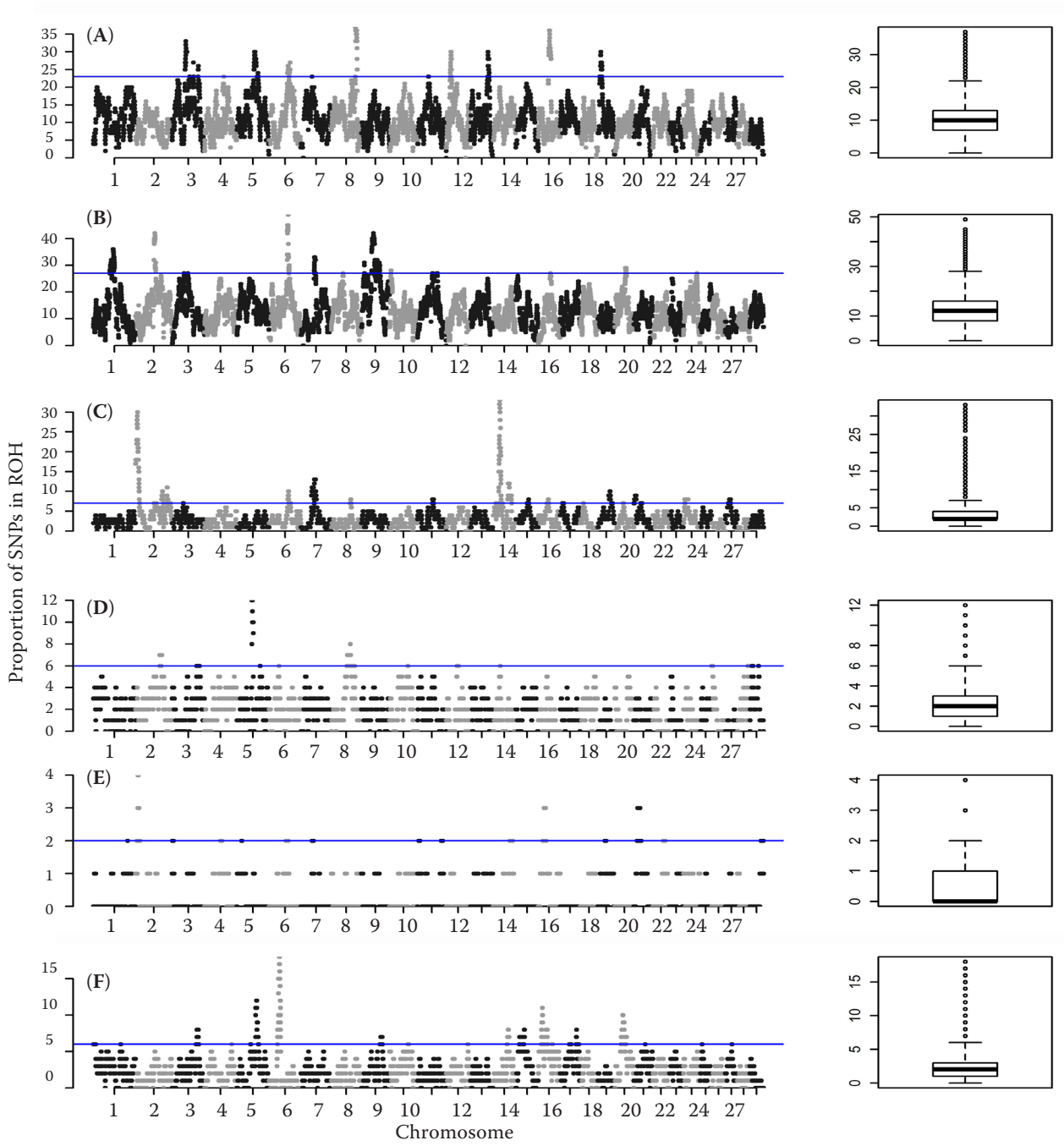

Figure 3. Occurrence of single nucleotide polymorphisms in runs of homozygosity (ROH) across the autosomal genome of analysed breeds: (**A**) Aberdeen Angus, (**B**) Hereford, (**C**) Limousin, (**D**) Charolais, (**E**) Piedmontese, (**F**) Romagnola

SNPs = single nucleotide polymorphisms

breed. A detailed description of the identified regions with an extreme frequency of ROH, including the number of genes and QTL regions, is given in Table 2.

In the Aberdeen Angus genome, nine selection signals were observed on chromosomes 3, 5, 6, 8, 12, 13, 16, and 19, covering overall 58.34 Mbp of the autosomal genome. The strongest signals were found on BTA8 and BTA16. In total, 866 protein-coding genes were identified directly in detected regions. Concerning meat production, these were e.g. the genes *IGF1* (BTA5), *GHRH* (BTA13), *MYOC* (BTA16) and *MYO19* (BTA19). The *IGF1* (insulin-like growth factor 1) gene is considered to be a candidate gene for several production traits, suitable also for MAS in cattle. Several studies confirmed its importance

https://doi.org/10.17221/226/2019-CJAS

Table 2. Autosomal regions under selection pressure identified based on runs of homozygosity segments distribution in the genome

Breed	BTA	Start (Mb)	End (Mb)	Region length (Mb)	Genes *n*	QTL in region
Aberdeen Angus	3	48.313	55.538	7.225	88	marbling score, milk yield, fat content
	5	64.190	72.618	8.428	99	dressing percentage, carcass weight, birth weight, LMA, twins frequency, FSH concentration
	6	69.027	78.477	9.450	74	growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content
	8	90.844	97.835	6.991	74	
	12	15.688	20.510	4.822	63	milk production, protein content, high loin area
	13	65.575	70.195	4.620	65	
	16	38.629	47.971	9.343	147	carcass weight
	19	7.731	11.999	4.268	87	EBV for backfat thickness, average daily gain
		12.669	15.864	3.194	74	dressing percentage
Hereford	1	61.936	68.605	6.668	76	carcass weight, dressing percentage
	2	68.691	75.724	7.033	54	marbling score, length of productive life
	6	67.240	75.137	7.898	88	growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content
	7	49.465	54.774	5.309	132	
	9	10.570	11.733	1.163	6	
Limousin	2	0.190	13.411	13.221	103	milk yield, carcass weight
	6	70.237	72.837	2.601	29	growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content
	7	39.530	46.478	6.947	269	intensity of ovulation, SCS
	14	21.934	30.595	8.661	67	average daily gain, milk yield, fat and protein content, IMF
Charolaise	2	96.021	97.795	1.774	29	marbling score, carcass weight
	5	56.375	61.729	5.353	219	dressing percentage, carcass weight, birth weight, LMA, EBV for backfat thickness, twins frequency
	8	68.660	70.253	1.593	23	birth difficulty (maternal effect), dystocia (maternal effect)
Piedmontese	2	8.000	10.618	2.618	14	marbling score, carcass weight
Romagnola	3	92.304	98.252	5.948	71	marbling score, milk yield
	5	70.200	78.116	7.916	102	dressing percentage, carcass weight, birth weight, LMA, twins frequency, FSH concentration
	6	31.354	42.446	11.092	52	birth weight, strength, power, LMA, average daily gain, milk yield, fat and protein content
	14	54.324	60.682	6.358	28	carcass weight, high loin area, birth weight
	15	32.862	33.310	0.449	5	
	16	12.131	16.454	4.323	22	marbling score
	17	61.118	62.291	1.173	8	
	20	26.303	35.684	9.382	56	milk yield, fat and protein content

BTA = *Bos taurus* autosome, QTL = quantitative trait locus, IMF = intramuscular fat, LMA = *Longissimus thoracis* muscle area, EBV = estimated breeding value, SCS = somatic cell score

https://doi.org/10.17221/226/2019-CJAS

mainly in connection with the growth intensity of beef cattle (Yurnalis et al. 2017). Similarly, the *GHRH* (growth hormone-releasing hormone) gene, which is secreted in the hypothalamus and stimulates the activity of growth hormone released from the thyroid gland, is associated with growth intensity and body weight. A significant effect of selection for this gene was also confirmed by Rothammer et al. (2013). Besides, the *KIT* and *KDR* genes were found in the region on BTA6 similarly like in the previous approach.

In Hereford cattle, the strongest signals were found on BTA1, BTA2, BTA6, BTA7 and BTA9. These regions covered 28.07 Mbp of the autosomal Hereford genome. The strongest selection signal was found on BTA6. Like in the case of Aberdeen Angus, genes controlling coat colour (*KIT* and *KDR*) were found in this region. Overall, 455 protein-coding genes and several QTLs for carcass traits, dressing percentage, marbling score or longevity were identified across all detected signals.

In the Limousin genome, four selection signals (BTA2, BTA6, BTA7 and BTA14) covering 31.43 Mbp of the genome were detected. However, the regions under the most intense selection pressure were located on chromosomes 2 and 14. This confirms previous studies in beef cattle, which reported a strong impact of selection mainly in the regions on chromosome 2 containing genes for intramuscular fat content and feed conversion (Qanbari et al. 2011) and in the centromeric region of BTA14 controlling marbling (Wiener et al. 2011). Besides, on BTA2 the *MSTN* gene was identified, which is responsible for double muscling in cattle (Fiems 2012).

In the Charolais genome, three regions under strong selection pressure were detected on BTA2, BTA5 and BTA8 (in total 8.72 Mbp). The strongest signal was shown by BTA5. In total 56 genes and various QTLs were identified in these regions, including those for dressing percentage, carcass weight, birth weight, *longissimus thoracis* muscle area, estimated breeding value for backfat thickness and twin frequency. For example, the *STAC3*, *MYO1A* and *MYL6* genes associated with skeleton muscle formation (Zhang et al. 2014b) were located in the selection signal regions.

In the Piedmontese breed, only a low effect of selection on the genome structure was observed. Based on previous results this trend was expected. The relatively weak signal was identified on BTA2 (2.16 Mbp) in the genomic region including 52 genes and QTLs for milk production and carcass weight.

In the genome of Romagnola cattle, the analysis proved that eight regions were significantly affected by selection. These regions were located on BTA3, BTA5, BTA6, BTA14, BTA15, BTA16, BTA17 and BTA20 and covered 46.64 Mbp of the autosomal genome. The strongest signal was observed on BTA5 and BTA6. Overall 853 protein-encoding genes associated with milk (*ABCG2*, *SPP1*) or meat production (*MYH9*, *GHR*) and reproduction (*PPARGC1A*) were found directly in the selection signals (Table 2). Perhaps the most important gene in terms of cattle performance is the gene encoding the growth hormone receptor (*GHR*). Growth hormone (GH) and its receptor are involved in the regulation of the metabolism of nutrients (e.g. carbohydrates, lipids, proteins and minerals). It has been revealed that

Table 3. Autosomal regions under selection pressure identified based on the variability in linkage disequilibrium

Breed	BTA	Start (Mb)	End (Mb)	Region length (Mb)	Genes *n*	QTL in region
Aberdeen Angus	13	62.825	65.859	3.034	82	dressing percentage
Hereford	7	47.201	49.331	2.130	39	
	13	54.208	55.559	1.350	70	dressing percentage
Limousin	26	19.088	22.122	3.034	47	dressing percentage, IMF, milk yield
Charolaise	20	32.621	37.325	4.704	39	
Piedmontese	7	48.308	50.202	1.894	53	
	20	33.418	35.197	1.778	7	milk production, fat and protein content
	26	20.572	21.075	0.503	11	dressing percentage, IMF, milk yield
Romagnola	6	38.791	41.800	3.009	25	birth weight, growth, strength, milk yield, fat and protein content

BTA = *Bos taurus* autosome, QTL = quantitative trait locus, IMF = intramuscular fat

https://doi.org/10.17221/226/2019-CJAS

GH is responsible mainly for tissue growth and lipid metabolism, simultaneously affecting lactation and normal body growth (Rothammer et al. 2013).

Variability in linkage disequilibrium. This approach was based on an assumption that the regions affected by selection show lower genetic

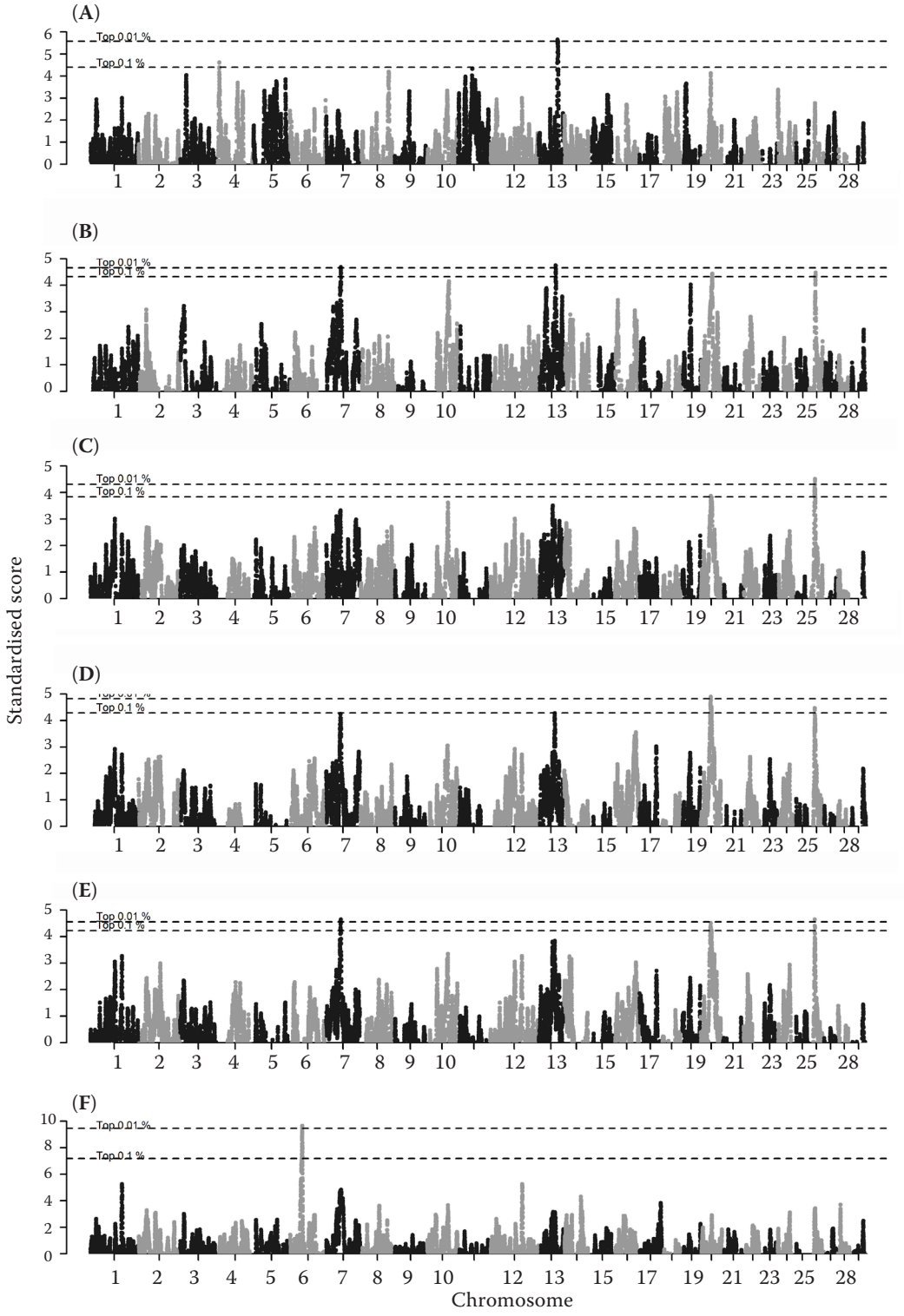

Figure 4. Distribution of variability in linkage disequilibrium (VarLD) score in the genome of analysed breeds: (**A**) Holstein/Aberdeen Angus, (**B**) Holstein/Hereford, (**C**) Holstein/Limousin, (**D**) Holstein/Charolais, (**E**) Holstein/Piedmontese, (**F**) Holstein/Romagnola

https://doi.org/10.17221/226/2019-CJAS

variability and a specific level of local LD between SNP markers. For a comparison of variation in LD, the Holstein cattle were used. A significant variability in LD was observed for all breeds, and the differences in the production type of breeds were manifested by the expression of specific genomic regions (Table 3). Figure 4 shows the standardised varLD score for all breeds studied.

Overall, five regions showing a significant level of variability in LD were found in the genome of analysed breeds. These regions were located on chromosomes 7, 6, 13, 20 and 26. Signals detected on BTA7, BTA13, BTA20 and BTA26 overlapped across the breeds (Table 3, Figure 5). The genomic region on BTA7 identified in Hereford and Piedmontese cattle contained 54 protein-coding genes, e.g. the *CAST* gene encoding calpastatin. Calpastatin, an inhibitor of the enzyme calpain, is an endogenous enzyme described as Ca-dependent proteinase. Calpastatin and calpain are activated in muscles of live animals as well as after slaughter during meat tenderisation when calpastatin inhibits calpain. In previous studies, a strong association between the *CAST* gene polymorphism and meat tenderness was found, especially for SNP CAST:c.155C>T (Barendse et al. 2007).

In Aberdeen Angus and Hereford breeds, a total of 82 genes were identified in the region on chromosome 13. Of these, the family of *LBP* genes (isoforms *BPIFA2A*, *BPIFA2C*, *BPIFA2B*, *BPIFA3*, *BPIFA1*, *BPIFB1* and *BPIFB5*), important for the intramuscular profile of muscles (Berton et al. 2016), accounted for a relatively large part. A similar region was identified by Sorbolini et al. (2015) in the genome of Piedmontese and Marcchigiana cattle. This region also contained several QTLs associated with dressing percentage.

In the region on BTA20 (Charolais and Piedmontese), 39 genes and several QTLs mainly for dairy performance were located. In the region on BTA26

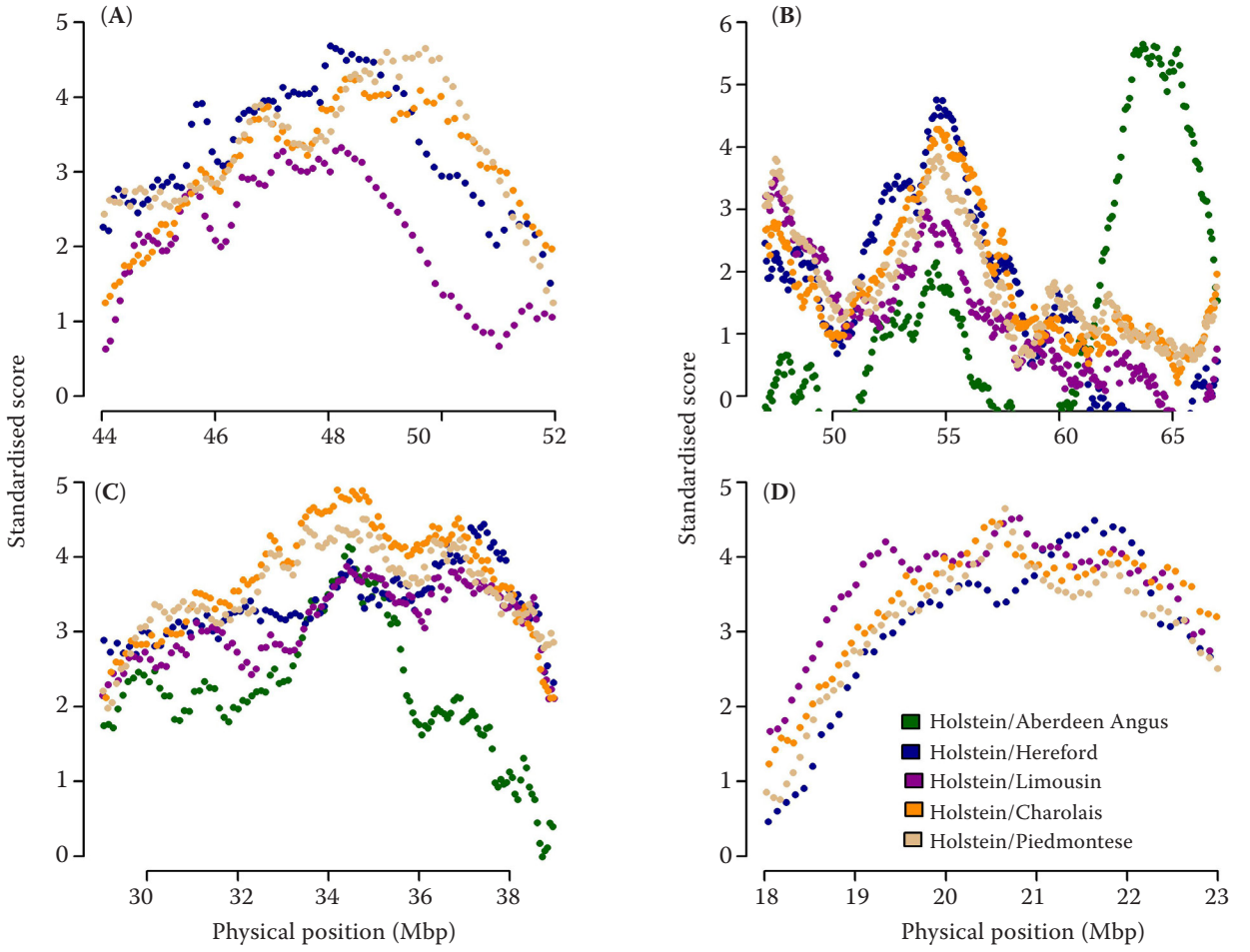

Figure 5. Normalised variability in linkage disequilibrium (VarLD) score for (**A**) BTA7, (**B**) BTA13, (**C**) BTA20 and (**D**) BTA26

https://doi.org/10.17221/226/2019-CJAS

(Limousin and Piedmontese), 47 genes were identified, as well as QTLs for dressing percentage, intramuscular fat content and milk production. In this region, the *SCD* gene (stearoyl-CoA desaturase) was also found, which is responsible for the conversion and regulation of the process by which saturated fatty acids are converted to monounsaturated in tissues. In mammalian tissues, the *SCD* is also referred to as delta 9-desaturase. One of the most commonly analysed polymorphisms for the *SCD* gene is the SNP SCD1.878 G> A, which is associated with intramuscular fat content in cattle (Li et al. 2013).

Only in the Romagnola genome, selection signatures on BTA6 were observed. Besides QTLs associated with meat performance, QTLs for milk production and fat and protein content were detected. Since the Romagnola breed was originally multi-purpose, the occurrence of such selection signals in the genome was expected. Several authors revealed that crossing of dairy and beef breeds could be beneficial in many cases, especially when beef quality is considered (Coleman et al. 2016).

CONCLUSION

Although the distribution of selection signals in the genome differs depending on the breed and the approach applied, several common signals were found on chromosomes 2, 6, 7, 13 and 20. These regions were located directly or very close to the sequences of genes controlling coat colour (*KIT* and *KDR*), muscle development (*GDF9*, *GHRH*, *GHR*), double muscling (*MSTN*), meat tenderness (*CAST*) and intramuscular fat content (*SCD*). Thus, the results confirmed that the selection pressure within the genome of the studied breeds was concentrated mainly in genomic regions controlling live weight, body frame, dressing percentage, carcass traits, meat quality, and longevity. In general, breeding programs for different cattle breeds overlap in some goals, so that approaches to the identification of selection signals can provide a new insight into the selection direction as well as defining slight differences in breeding programs. Simultaneously, specific regions for particular breeds were identified. A closer analysis of the genotypes of SNP markers located directly in the region of the gene sequences, as well as their involvement in MAS or genomic selection, could be beneficial for the genetic progress of those breeds in future.

REFERENCES

Barendse W., Harrison B.E., Hawken R.J., Ferguson D.M., Thompson J.M., Thomas M.B., Bunch R.J. (2007): Epistasis between calpain 1 and its inhibitor calpastatin within breeds of cattle. Genetics, 176, 2601–2610.

Bellinge R.H., Liberles D.A., Iaschi S.P., O'Brien P.A., Tay G.K. (2005): Myostatin and its implications on animal breeding: A review. Animal Genetics, 36, 1–6.

Berton M.P., Fonseca L.F., Gimenez D.F., Utembergue B.L., Cesar A.S., Coutinho L.L., de Lemos M.V., Aboujaoude C., Pereira A.S., Silva R.M., Stafuzza N.B., Feitosa F.L., et al. (2016): Gene expression profile of intramuscular muscle in Nellore cattle with extreme values of fatty acid. BMC Genomics, 17, 972.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J. (2015): Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4, 7.

Coleman L.W., Hickson R.E., Schreurs N.M., Martin N.P., Kenyon P.R., Lopez-Villalobos N., Morris S.T. (2016): Carcass characteristics and meat quality of Hereford sired steers born to beef-cross-dairy and Angus breeding cows. Meat Science, 121, 403–408.

Dorshorst B., Henegar C., Liao X., Sallman Almen M., Rubin C.J., Ito S., Wakamatsu K., Stothard P., Van Doormaal B., Plastow G., Barsh G.S., Andersson L. (2015): Dominant red coat color in Holstein cattle is associated with a missense mutation in the coatomer protein complex, subunit alpha (COPA) gene. PLoS ONE, 10, e0128969.

Fiems L.O. (2012): Double muscling in cattle: genes, husbandry, carcasses and meat. Animals, 2, 472–506.

Fontanesi L., Scotti E., Russo V. (2010): 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. Italian Journal of Animal Science, 9, e42.

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

Li X., Ekerljung M., Lundstrom K., Lunden A. (2013): Association of polymorphisms at DGAT1, leptin, SCD1, CAPN1 and CAST genes with color, marbling and water holding capacity in meat from beef cattle populations in Sweden. Meat Science, 94, 153–158.

McTavish E.J., Decker J.E., Schnabel R.D., Taylor J.F., Hillis D.M. (2013): New World cattle show ancestry from multiple independent domestication events. Proceedings of the National Academy of Sciences of the United States of America, 110, 1398–1406.

https://doi.org/10.17221/226/2019-CJAS

Moravcikova N., Simcic M., Meszaros G., Solkner J., Kukuckova V., Vlcek M., Trakovicka A., Kadlecik O., Kasarda R. (2018): Genomic response to natural selection within Alpine cattle breeds. Czech Journal of Animal Science, 63, 136–143.

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G., Simianer H. (2011): Application of site and haplotype-frequency based approach for detecting selection signatures in cattle. BMC Genomics, 12, 318.

Randhawa I.A., Khatkar M.S., Thomson P.C., Raadsma H.W. (2016): A meta-assembly of selection signatures in cattle. PLoS ONE, 11, e0153013.

Rothammer S., Seichter D., Forster M., Medugorac I. (2013): A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics, 14, 908.

Sabeti P.C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., Xie X., Byrne E.H., McCarroll S.A., Gaudet R., Schaffner S.F., Lander E.S., et al. (2007): Genome-wide detection and characterization of positive selection in human populations. Nature, 449, 913–918.

Salem M.M.I., Thompson G., Chen S., Beja-Pereira A., Carvalheira J. (2018): Linkage disequilibrium and haplotype block structure in Portuguese Holstein cattle. Czech Journal of Animal Science, 63, 61–69.

Sharifiyazdi H., Mirzaei A., Ghanaatian Z. (2018): Characterization of polymorphism in the FSH receptor gene and its impact on some reproductive indices in dairy cows. Animal Reproduction Science, 188, 45–50.

Sorbolini S., Marras G., Gaspa G., Dimauro C., Cellesi M., Valentini A., Macciotta N.P. (2015): Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. Genetics Selection Evolution, 47, 52.

Tanq K.Q., Yang W.C., Li S.J., Yang L.G. (2013): Polymorphisms of the bovine growth differentiation factor 9 gene associated with superovulation performance in Chinese Holstein cows. Genetics and Molecular Research, 12, 390–399.

Teo Y.Y., Fry A.E., Bhattacharya K., Small K.S, Kwiatkowski D.P., Clark T.G. (2009): Genome-wide comparisons of variation in linkage disequilibrium. Genome Research, 19, 1849–1860.

The Bovine HapMap Consortium (2009): Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science, 324, 528–532.

Utsunomiya Y.T., O'Brien A.M., Sonstegard T.S., Solkner J., Garcia J.F. (2015): Genomic data as the "hitchhiker's guide" to cattle adaptation: Tracking the milestones of past selection in the bovine genome. Frontiers in Genetics, 6, 1–13.

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.

Wiener P., Gutierrez-Gil B. (2009): Assessment of selection mapping near the myostatin gene (GDF-8) in cattle. Animal Genetics, 40, 598–608.

Wiener P., Edriss M.A., Williams J.L., Waddington D., Law A., Woolliams J.A., Gutierrez-Gil B. (2011): Information content in genome-wide scans: Concordance between patterns of genetic differentiation and linkage mapping associations. BMC Genomics, 12, 65.

Yurnalis, Arnim, Putra D.E. (2017): Polymorphism of insulin-like growth factor 1 gene (IGF1/TasI, IGF1/SnaBI, IGF1/RsaI) and the association with daily gain of Pesisir cattle local breed from West Sumatera, Indonesia. Pakistan Journal of Biological Sciences, 20, 210–216.

Zhang Y., Yang M., Li C., Xu Y., Sun J., Lei C., Lan X., Zhang C., Chen H. (2014a): Identification and genetic effect of a variable duplication in the promoter region of the cattle ADIPOQ gene. Animal Genetics, 45, 171–179.

Zhang Y., Cong X., Wang A., Jiang H. (2014b): Identification of the STAC3 gene as a skeletal muscle-specifically expressed gene and a novel regulator of satellite cell differentiation in cattle. Journal of Animal Science, 92, 3284–3290.

Zheng L., Zhang G.M., Dong Y.P., Wen Y.F., Dong D., Lei C.Z., Qi X.L., Chen H., Huo L.J., Huang Y.Z. (2019): Genetic variant of MYLK4 gene and its association with growth traits in Chinese cattle. Animal Biotechnology, 30, 30–35.

Received: 2019–10–30
Accepted: 2019–12–12