

Genome-wide characterisation of regions under intense selection based on runs of homozygosity in Charolais cattle

Barbora Olšanská*, Radovan Kasarda, Kristína Lehocká, Nina Moravčíková

Slovak University of Agriculture, Department of Genetics and Animal Breeding Biology, Nitra, Slovakia



Licensed under a Creative Commons Attribution 4.0 International License



In this study, 68 genotyped purebred cows and bulls of Charolais cattle were used to determine runs of homozygosity (*ROH*) for evaluation of selection effect on the genome structure. *ROH* analysis was performed for 41,153 SNPs, and homozygous segments considering a minimum of 15 consecutive SNPs. The aim was to identify if regions of the genome with a high frequency of SNPs in *ROH* include signatures of selection. The most significant outlier SNPs were found on BTA2, 5, 7 and 19 (11 regions), with a sum of *ROH* length of 15.85 Mb. These regions contained genes included in various biological processes associated with the functioning of the immune system, growth, reproduction and metabolism. Various quantitative traits loci (QTLs) which affected the characteristics of meat production and reproduction have been identified in these regions. Overall, obtained results suggest that the Charolais cattle genome includes selection signatures reflecting the improvement of meat production and reproduction in accordance to breeding objectives.

Keywords: artificial selection, autozygosity, beef, bovine, candidate genes

1 Introduction

The Charolais cattle is the popular beef breed worldwide. Due to the number of animals, Charolais is one of the most numerous beef breeds in Slovakia. In 2019, 10,225 cows and 283 bulls were recorded. The Charolais cattle originated from France and is characterised by a robust and muscular body, excellent feed conversion rate and calving ease. The main breeding goal included the effort to create a population of modern animals with excellent meat production in combination with maintaining good adaptability to the natural environment, excellent maternal characteristics, good ability of grazing and hornless (Jahuey-Martínet et al., 2019; The Breeding Services of the Slovak Republic, s.e., 2019; Pomichal, 2009).

Selection left footprints in the genome, which were characterised by high genetic differentiation between breeds or an evident decrease in genetic diversity in genomic regions associated with traits that undergone directional selection. The identification of selection signals involved in phenotypic variations was important for understanding the evolutionary processes and mechanisms that underlie properties that have been subjected to natural or artificial selection (Mastrangelo et al., 2020).

The purpose of the analysis of selection signals is to identify genomic regions or loci that showed deviations from neutrality. Genomic regions that have been affected by the intensive selection showed reduced genetic diversity and high homozygosity, and these regions can be called runs of homozygosity (*ROH*) islands. Current genomic tools have enabled the genotyping of thousands of polymorphisms and thus made it possible to identify long stretches of homozygous genotypes in the genome. Two identical inherited haplotypes from the parents create homozygous segments in the genome, i.e. *ROH* (Onzima et al., 2018; Szmatoła et al., 2020; Shi et al., 2020; Mastrangelo et al., 2019; Purfield et al., 2017).

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

Genome-wide analysis of *ROH* may be performed to understand population history, calculate genomic inbreeding, decipher the genetic architecture of complex traits and diseases, as well as to identify genes associated with economically important traits (Dixit et al., 2020).

The purpose of this study was to characterize unique regions potentially under selection and identify genes and metabolic pathways associated with traits of economic interest in Charolais cattle.

2 Material and methods

A total of 45 purebred cows and 23 bulls of Charolais cattle were genotyped by International Dairy and Beef Chip. Quality control of genotyped data was performed in PLINK v1.9 software (Chang et al., 2015) according to set up of Moravčíková et al. (2018): call rate for SNPs and for individuals > 90 %, MAF > 0.01.

Runs of homozygosity were computed using detectRUNS package in R software environment (Biscarini et al., 2019). Criteria used for defining *ROHs* were as follows: minimum number of homozygous SNPs in window = 15, maximum number of heterozygous SNPs = 0, maximum number of missing genotypes = 1. The frequency of SNPs in *ROH* longer than 4 Mb were calculated. The outlier SNPs denoted as genomic regions under intense selection were determined by the upper quartile of boxplot and the threshold was set to 8% SNPs in *ROH*. For each region affected by selection, a list of genes was created using Genome Data Viewer by assembly ARS-UCD1.2 and metabolic pathways were identified by web-based toolkit WebGestalt (<http://www.webgestalt.org/>).

3 Results and discussion

After applying the quality control, the final dataset included 41,153 SNPs with an overall length of genome of 2,503 Mb and average distance between adjacent loci of 60.87 ± 63.45 kb. The minimum distance between SNPs was 0.028 kb, and the maximal distance was 2,141.34 kb. Selection signals were identified for all autosomes across the genome and for the subsequent analysis four chromosomes where selection focused on traits of economic importance were chosen (Fig 1). Analysis of homozygous segments within the Charolais genome allowed to determine genomic regions with a high frequency of SNPs in *ROH*. The average length of *ROH* was 5.34 ± 4.43 Mb, and overall length of *ROH* was 15.85 Mb. In this study the most significant outlier SNPs were located in 11 regions in 4 chromosomes (BTA2, 5, 7, and 19).

Compared to our results Szmatoła et al. (2019) reported five affected by directional selection regions for Charolais cattle and they observed two regions in approximately the same position (BTA5 53.26 - 62.18 Mb and BTA7 28.18 - 31.97 Mb). Mészáros et al. (2015) identified the strongest selection signal in the position from 36 to 41 Mb for Tyrol Grey on BTA6. Moravčíková et al. (2019) reported three genomic regions under intense selection on BTA2, 5 and 8 for Charolais cattle and compared to our results, we found the same genomic region affected by selection on chromosome 5 (56.38 – 61.73 Mb). Summary of the regions under selection is listed in Table 1.

Detailed examination of identified signals showed that affected genomic regions were subject to intensive selection for the meat production, reproduction and milk components. The strongest selection signal was observed on BTA5 in the area of 74 protein-coding genes, and various QTLs which affected mainly reproduction and milk composition. The genes *HOXC4*, *HOXC5*, *HOXC6*, *HOXC8*, *RARG* of BTA5 (26.15 – 26.99 Mb) are involved in the skeletal system development (Gaudet et al., 2011). Within BTA2 (6.22 - 6.23 Mb), gene *MSTN*; the latter is known for a regulator of muscle growth factor (Trukhachev et al., 2015). The *FAF2* gene was identified in BTA7 (39.41 – 39.47 Mb). Based on Patel et al. (2017), this gene is associated with some reproductive function in cattle. Gene *NLRP3* involved in the innate immune system was also observed on BTA7 (42.18 – 42.23) (Mallikarjunappa et al., 2018). Gene *CA10* is exclusively expressed in the brain, and it is highly conserved across animal species, was found on BTA 19 (0.60 – 1.50 Mb) (Aspatwar et al., 2014).

4 Conclusions

This study provided information about potential candidate genes and QTL regions affected by directional selection in Charolais cattle. The results indicate several genomic regions affected by intensive selection. The formation of these regions may be based on adaptation to environmental conditions, such as immune system function, disease resistance, response to stimulus these regions also reflect efforts to improve meat production and reproduction traits such as calving ease. This study

provided proofs that some areas of the genome in Charolais have been affected by intensive long-term selection on meat-related traits and on specific traits of interest for beef cattle.

Table 1 Characterisation of autosomal regions affected by selection

CHR	Position (Mb)	QTL	Genes
2	5.88 - 7.56	luteal activity	<i>INPP1, HIBCH, MSTN, PMS1, ORMDL1, OSGEPL1, ANKAR, ASNSD1, SLC40A1, WDR75, COL5A2, MIR2917, COL3A1</i>
	9.04 - 11.49	luteal activity, marbling score	<i>ZSWIM2, MIR2351, ITGAV, ZC3H15, FSIP2</i>
5	22.08 - 27.09	daughter pregnancy rate, milk fat yield, milk yield	<i>BTG1, EEA1, NUDT4, UBE2N, MRPL42, SOCS2, CRADD, CEP83, TMCC3, METAP2, USP44, MUCL1, GLYCAM1, PPP1R1A, PDE1B, NCKAP1L, GTSF1, ITGA5, ZNF385A, GPR84, COPZ1, MIR148B, NFE2, CBX5, SMUG1, HOXC4, HOXC5, MIR615, HOXC6, HOXC8, HOXC9, HOXC11, HOXC12, HOXC13, PRR13, AMHR2, SP1, SP7, AAAS, PFDN5, ESPL1, MFSD5, RARG, ITGB7, ZNF740, CSAD, SOAT2, IGF2BP6, SPRYD3, TNS2, EIF4B, KRT18, KRT8, KRT78, KRT79, KRT4, KRT3</i>
	53.26 - 61.36	milk yield, loin strength, milk mid-infrared spectra, tenderness score, milk protein percentage, abomasum displacement, milk fat percentage	<i>SLC16A7, R3HDM2, NTN4, OS9, LRP1, CS, CDK17, ELK3</i>
	73.63 - 82.04	daughter pregnancy rate, milk fat yield, conception rate, milk butyric acid content	<i>RBFOX2, MYH9, CANG2, SYT10, PKP2, FGD4, BICD1, AMN1, DENND5B, TMTC1</i>
7	30.51 - 34.01	milk beta-casein percentage, scrotal circumference, lean meat yield, reproductive efficiency	<i>ZNF608, CSNK1G3, CEP120, PRDM6, PPIC, SNX24, SNX2, SNCAIP, SRFBP1, FTMT, LOX</i>
	39.22 - 46.35	milk fat yield, bone quality, calving ease (maternal), stillbirth (maternal), foot angle, feet and leg conformation, teat placement, udder placement	<i>COMMD10, UIMC1, FAF2, COL23A1, NLRP3</i>
19	0.47 - 4.70	carcass weight, average daily gain, subcutaneous fat, body weight (yearling), scrotal circumference, stature, insemination per conception, somatic cell score, rump angle, fat thickness at the 12th rib	<i>CA10, KIF2B</i>
	36.63 - 38.47		<i>ABCC3, CACNA1G, ACSF2, ITGA3, KAT7, SPOP, B4GALNT2, IGF2BP1</i>
	41.92 - 57.77	milk fat percentage, myristic acid content, somatic cell score, myristic acid content, myristoleic acid content, palmitoleic acid content, milk unglycosylated kappa-casein percentage	<i>NSF, TLK2, TANC2, TBCD, RPTOR, RBFOX3</i>

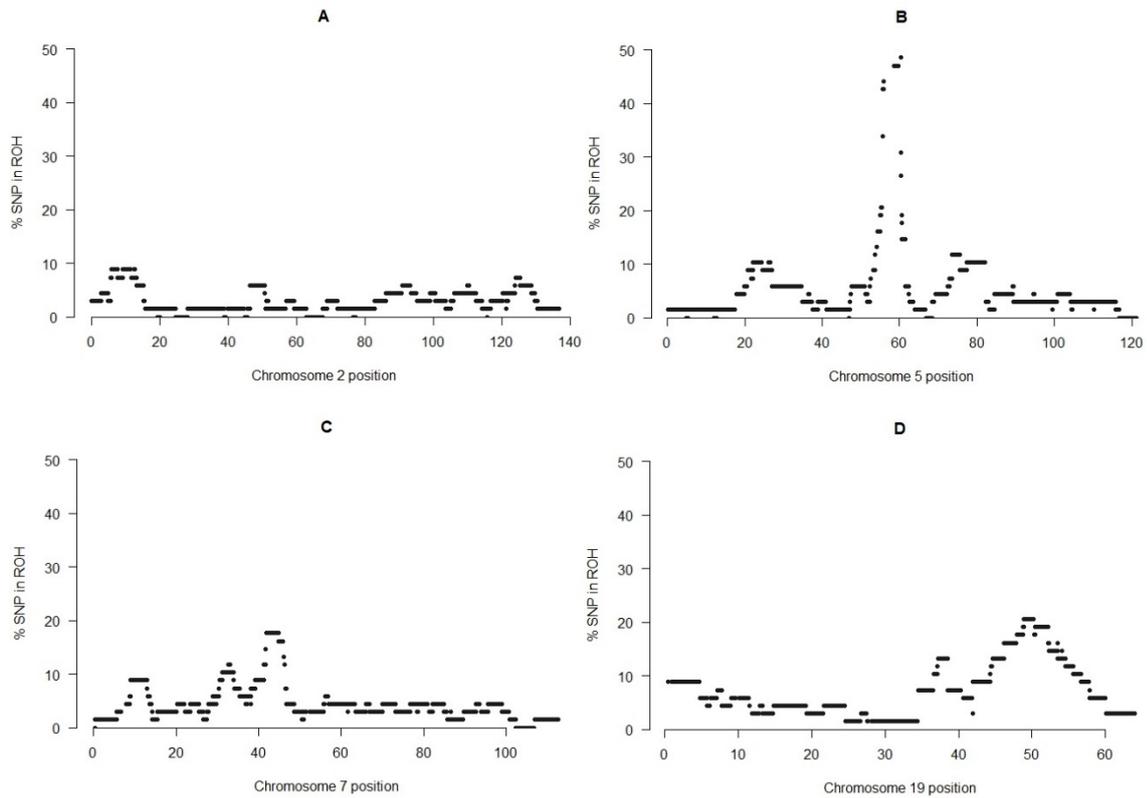


Figure 1 The occurrence of SNP in *ROH* (%) across the autosomes BTA2 (A), BTA5 (B), BTA7 (C) and BTA19 (D)

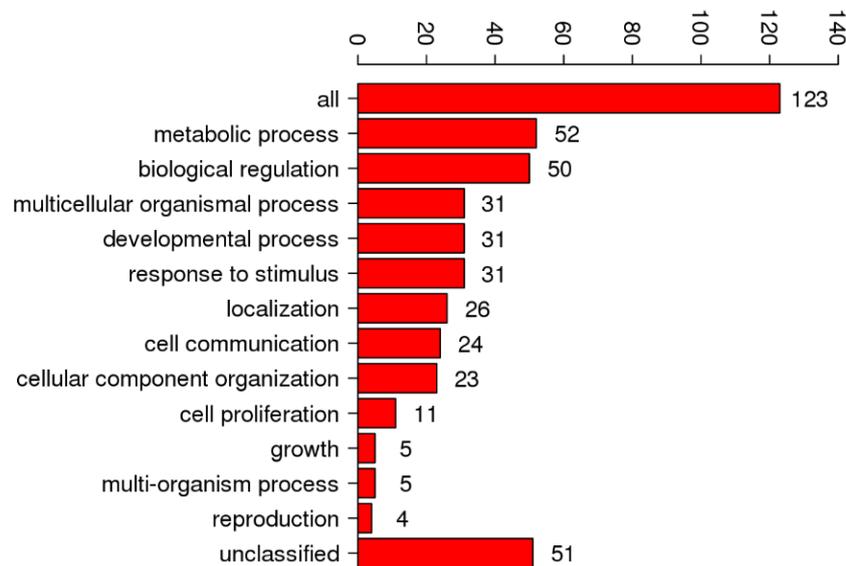


Figure 2 Number of genes overlapped with the annotated genes in the GO Slim terms from biological process across the regions with high occurrence of *ROH*

Acknowledgments

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

References

- Aspatwar, A. et al. (2014). Carbonic anhydrase related proteins: molecular biology and evolution. *Subcell. Biochem.* 75: 135–156.
- Biscarini, F. et al. (2019). *detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes*. The R Project. Retrieved June 8, 2019 from <https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html>
- Dixit, S. P. et al. (2020). Genome-Wide Runs of Homozygosity Revealed Selection Signatures in *Bos indicus*. *Front. Genet.* <https://doi.org/10.3389/fgene.2020.00092>
- Gaudet, P. et al. (2011). Phylogenetic-based Propagation of Functional Annotations Within the Gene Ontology Consortium. *Brief Bioinform.* 12(5): 446-62. <https://doi.org/10.1093/bib/bbr042>
- Chang, C.C. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>
- Jahuey-Martinez, F. L. et al. (2019). Signatures of selection in Charolais beef cattle identified by genome-wide analysis. *Journal of Animal Breeding and Genetics*, 136(5). <https://doi.org/10.1111/jbg.12399>
- Mallikarjunappa, S. et al. (2018) Short communication: Uncovering quantitative trait loci associated with resistance to *Mycobacterium avium* ssp. *paratuberculosis* infection in Holstein cattle using a high-density single nucleotide polymorphism panel, *Journal of Dairy Science*, 101(8):7280-7286. <https://doi.org/10.3168/jds.2018-14388>
- Marras, G. et al. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.*, 46, 110–121. <https://doi.org/10.1111/age.12259>
- Mastrangelo, S. (2020). Genome-wide detection of signatures of selection in three Valdostana cattle populations. *Journal of Animal Breeding and Genetics.* <https://doi.org/10.1111/jbg.12476>
- Mastrangelo, S. et al. (2019). Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. In *PLoS ONE* 14(6): e0209632. <https://doi.org/10.1371/journal.pone.0209632>
- McTavish E.J. (2013). New world cattle show ancestry from multiple independent domestication events. *PNAS*, 110(15): 1398–1406. <https://doi.org/10.1073/pnas.1303367110>
- Mészáros, G. et al. (2015) Genomic analysis for managing small and endangered populations: A case study in Tyrol Grey cattle. *Front. Genet.* 13(6):173. <https://doi.org/10.3389/fgene.2015.00173>
- Moravčíková, N. et al. (2019). Analysis of selection signatures in the beef cattle genome. *Czech Journal of Animal Science*, 64(12): 491–503. <https://doi.org/10.17221/226/2019-CJAS>
- Moravčíková, N. et al.(2018). Genomic Response to Natural Selection within Alpine Cattle Breeds. *Czech J. Anim. Sci.*, 63(4):136-143. <https://doi.org/10.17221/62/2017-CJAS>
- Onzima, R. B. et al. (2018). Genome-Wide Characterisation of Selection Signatures and Runs of Homozygosity in Ugandan Goat Breeds. *Front. Genet.* <https://doi.org/10.3389/fgene.2018.00318>
- Patel, A. B. et al. (2017). Identification of single nucleotide polymorphism from Indian *Bubalus bubalis* through targeted sequence capture. *Current Science*, 112(6): 1230. ISSN 0011-3891
- Pomichal. (2009) *The breeding program of Charolais cattle*. The Slovak Beef Breeders Association, cooperative. Retrieved June 8, 2019 from <http://www.zchmd.eu/charolais>
- Purfield, D.C. et al. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.*, 13, 70. <https://doi.org/10.1186/1471-2156-13-70>
- Purfield, D.C. et al. (2017). The Distribution of Runs of Homozygosity and Selection Signatures in Six Commercial Meat Sheep Breeds. In *PLoS ONE* 12(5): e0176780. <https://doi.org/10.1371/journal.pone.0176780>
- Shi, L. et al. (2020). Estimation of inbreeding and identification of regions under heavy selection based on runs of homozygosity in a Large White pig population. *Journal of Animal Science and Biotechnology*, 11(46). <https://doi.org/10.1186/s40104-020-00447-0>
- Szmatola, T. et al. (2019). A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types. *Animals*, 9(12), <https://doi.org/10.3390/ani9121024>
- Szmatola, T. et al. (2020). Detection of runs of homozygosity in conserved and commercial pig breeds in Poland. *Journal of Animal Breeding and Genetics.* <https://doi.org/10.1111/jbg.12482>

The Breeding Services of the Slovak Republic, s.e. (2019). *The results of beef and suckling cows performance recording in Slovak Republic, Year 2019*. The Breeding Services of the Slovak Republic, s.e. Retrieved June 8, 2020 from <http://test.plis.sk/volne/rocnkamagazin/rocnka.aspx?id=kump2019>

Trukhachev, V. et al. (2015). Myostatin gene (MSTN) polymorphism with a negative effect on meat productivity in Dzhalginsky Merino sheep breed. In *J.BioSci. Biotechnol.* 4(2):191-199