

EVIDENCE OF SELECTIVE SWEEPS THROUGH HAPLOTYPE STRUCTURE OF PINZGAU CATTLE

Radovan KASARDA ¹, Nina MORAVČÍKOVÁ ², Veronika KUKUČKOVÁ ³,
Ondrej KADLEČÍK ⁴, Anna TRAKOVICKÁ ⁵, Gábor MÉSZÁROS ⁶

ABSTRACT

The aim of this study was to identify the genomic regions potentially under positive selection through the determination of integrated Haplotype Homozygosity Score in dual-purpose Pinzgau cattle. The genomic data from in total 37 Slovak and 105 Austrian purebred bulls were obtained using Illumina high-density SNP panels. After applying the quality control the final dataset composed of overall 27,895 autosomal loci with SNP density at the level one locus per 89.31 kb. The single-site iHS values have been scored for each SNP and subsequently averaged into the non-overlapping 500 kb windows across the genome. Totally, 27,293 iHS values across 4968 sliding windows were tested for the evidence of selection signatures. The most significant iHS values were detected for 6 genomic regions on BTA 2, 4, 5, 7, 8 and 22. The iHS value was in average 0.03 and the highest score of iHS (2.09) has been found in region on BTA 5. Our results provide initial insight and basis for future studies of genome-wide associations and impact of artificial selection in Slovak Pinzgau cattle.

Key words: cattle, breeds, Pinzgau, genetics, selection signals, high-density SNP arrays, iHS score

1 INTRODUCTION

The Pinzgau cattle in Europe traditionally belongs to the dual-purpose cattle breeds with combined milk and meat production (Pavlík *et al.*, 2013; Kadlečík *et al.*, 2004). The nucleus of the breed is based upon two populations registered in Herd books of Austria and Slovakia. In Slovak republic the Pinzgau cattle is bred for more than 150 years. It was imported from the Austrian Alps and therefore the two populations are well connected (Kadlečík *et al.*, 2011). The breeding management is specific in Slovakia however, where the genetic reserve is bred under intensive farming conditions in sub-mountainous regions. The nucleus of the breed is formed from

almost 50 % (in Slovakia) and more than 65 % (in Austria) of animals in suckling-cow system (Kasarda *et al.*, 2016). The breeders of Pinzgau cattle in Slovakia and also in Austria are interested in the future development of the breed, when integrating both pedigree and genomic data (Kasarda *et al.*, 2009; Mészáros *et al.*, 2010; Kasarda *et al.*, 2015).

Generally, the formation of cattle breeds is the result of human selection, adaptation to different environments and demographic effects such as domestication, migration and selection, all contributing to the patterns of genetic diversity (Mancini *et al.*, 2014). Artificial selection in cattle has resulted in divergent breeds that are specialized for either milk or meat production or raised

1 Slovak University of Agriculture in Nitra, Faculty of Agrobiolgy and Food Resources, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 94976 Nitra, Slovakia, e-mail: radovan.kasarda@uniag.sk

2 Same address as 1, e-mail: nina.moravcikova@uniag.sk

3 Same address as 1, e-mail: veronika.sidlova@gmail.com

4 Same address as 1, e-mail: ondrej.kadlecik@uniag.sk

5 Same address as 1, e-mail: anna.trakovicka@uniag.sk

6 University of Natural Resources and Life Sciences, Division Livestock Science, Augasse 2-6, 1180 Vienna, Austria, e-mail: gabor.meszaros@boku.ac.at

as dual-purpose breeds. The selection strategies for traits of interest led to the increase of selection pressure on particular genomic regions that control traits of productive and economic importance (Zhao *et al.*, 2015). When selection acts on a mutation, it also affects linked loci and leaves so called signatures in the flanking chromosomal regions. If intensive selection pressure occurred only over a few generations, it is unlikely that recombination has an impact on haplotype structure, and thus it resulted in linkage disequilibrium patterns between the mutation and the neighbouring loci. The signal of such selection pressure on genes can be detected as a spectrum of allele frequencies among closely linked loci that is shifted towards extreme frequencies, an excess of homozygous genotypes, and a high frequency of long haplotypes (Sabeti *et al.*, 2002; Makina *et al.*, 2015; Zhao *et al.*, 2015). The understanding of these signatures of selection is of interest from the perspective of evolutionary biology. Identification of candidate genes within these regions may ultimately help to further improve these breeds through informed management of breeding programmes (Qanbari and Simianer, 2014).

The availability of high density SNP genotyping arrays opens up possibility of performing population genetic studies in cattle using thousands of markers distributed across the entire genome (Urbinati *et al.*, 2016). The selection signature regions can be detected via several approaches. The studies applied in detection of selective sweeps are mainly based on estimation of differences in allele frequencies by contrasting pair of breeds through F_{ST} index, extended regions of low diversity, the calculation of extended haplotype homozygosity (EHH), and variants of this statistics such as Relative Extended Haplotype Homozygosity (REHH), the long-range haplotype (LRH) test, or integrated Haplotype Homozygosity Score (iHS) (Gutiérrez-Gil *et al.*, 2015).

The aim of this study was to perform the genetic analysis in two Pinzgau cattle populations with Slovak and Austrian origin in order to determine genomic regions, potentially under selection pressure and to further describe these regions. In our study the integrated Haplotype Homozygosity Score has been considered as an indicator of the genome-wide pattern of positive selection.

2 MATERIAL AND METHODS

Genome wide single nucleotide polymorphism (SNP) data from 37 Slovak and 105 Austrian purebred Pinzgau bulls were used to detect selective sweeps. The genotyping was performed using Illumina high-density SNP panels (50K and 700K). Overall 42,444 SNPs common to both genotyping arrays were retained in reduced

panel of loci. Markers assigned to unmapped regions or with unknown chromosomal position according to the latest bovine genome assembly (Btau 4.0) and SNPs positioned to sex chromosomes were removed. The quality control of genotyping data were performed according to Purcell *et al.* (2007) to exclude any autosomal locus with call rate lower than 90 %, minor allele frequency lower than 0.05 and HWE limit of 0.0001. The final dataset has been composed of 27,895 autosomal loci with SNP density at the level one locus per 89.31 kb.

The evidence of positive selection was determined based on evaluation of integrated Haplotype Homozygosity Score. The analysis of selective sweeps using iHS statistics is based on calculation of haplotype frequencies as specified by Voight *et al.* (2006). The haplotypes were reconstructed for each autosome using default options in fastPHASE software according to Scheet and Stephens (2006). Because the iHS approach requires the information of the ancestral and derived allele status for each locus, the set of ancestral alleles was established according to Rocha *et al.* (2014). In iHS statistics each locus is treated as core SNP and the test begins with calculation of extended Haplotype Homozygosity (EHH) for each core SNP. For a given focal SNP, the iHS is defined as the standardized log-ratio of the iHH (integrated EHH) for the ancestral (iHH_a) and derived (iHH_d) allele:

$$iHS^{(s)} = \frac{\log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right) - E_{ps} \log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right)}{sd_{ps} \log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right)}$$

The expectation $E_{ps} \log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right)$ and standard deviation

$sd_{ps} \log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right)$ of the unstandardized score

$\log\left(\frac{iHH_a^{(s)}}{iHH_d^{(s)}}\right)$ are estimated from the empirical distribu-

tion at SNPs whose derived allele frequency p_s matches the frequency at the core SNPs. Since iHS is standardized using the genome-wide empirical distributions, it provides a measure of how unusual the haplotypes around a given SNP are, relative to the whole genome (Voight *et al.*, 2006). The iHS score has been computed using the R package “rehh” (Gautier and Vitalis, 2012). The single-site iHS values were averaged within non-overlapping windows of 500 kb across the genome. The observed major peaks defined regions with extreme iHS values based

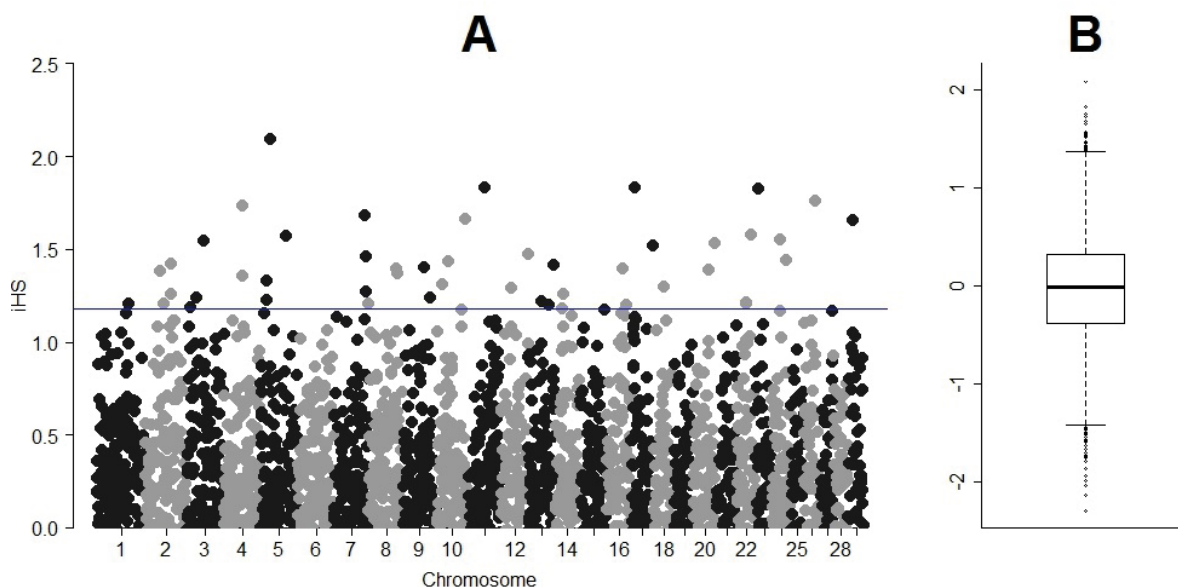


Figure 1: Distribution of Integrated Haplotype Score (iHS) averaged within 500 kb across the genome (A) and the boxplot of the iHS values (B)

on the outliers according to obtained boxplot distribution were evaluated as signals of the positive selection.

3 RESULTS AND DISCUSSION

The iHS approach introduced by Voight *et al.* (2006) was applied to determine the selective sweeps across Pinzgau populations. This approach is based on the ratio of the integrated EHH of two ancestral and derived core alleles. The ancestral alleles were defined based on genomic data of species considered to be a common founder of the *Bovinae* subfamily, in which the alleles were fixed at MAF equal to zero as demonstrated Utsunomiya *et al.* (2013) and Rocha *et al.* (2014). The single-site iHS values have been scored for each SNP and subsequently averaged into the non-overlapping 500 kb windows across the genome. The sliding windows size was adopted based on approach described by Qanbari *et al.* (2011). Overall 27,293 iHS values across 4968 sliding windows were tested for the evidence of selection pressure. Negative iHS values indicated long haplotypes carrying derived alleles and positive iHS values represented long haplotypes carrying ancestral alleles, and small or close to zero iHS values indicated similar rates of EHH decay for ancestral and derived SNP alleles (Urbinati *et al.*, 2016). However, an extreme positive iHS score has been considered as a candidate because the ancestral allele itself may be the target of selection (Qanbari and Simianer, 2014). The signals of recent selection were defined as genomic regions with positive iHS values based on the outliers according

to boxplot distribution (Fig. 1B). Thus, the genomic regions were considered as recently selected when the iHS score of multiple loci located within 0.5 Mb was greater than 1.18. Table 1 shows the summary of autosomal genomic regions displaying the strongest iHS signals. Total number of SNPs that can be regarded as loci under selection was lower in comparison to the study of Urbinati *et al.* (2016), Makina *et al.* (2015) or Zhao *et al.* (2015). The most significant iHS value exceeding the threshold 1.18 was detected across genome for 6 sliding windows (Table 1). Figure 1 shows the genome distribution of iHS values averaged within 500 kb against the genomic position. The highest score of iHS (2.09) has been found in the region on chromosome 5. Between the selective sweeps, no major overlap has been identified. The obtained genomic regions did not show strong signal of selection possibly due to close relatedness among both Pinzgau populations.

Table 1: Summary of genomic regions with the strongest iHS signals

BTA	Start (Mb)	End (Mb)	Average iHS value	No. of genes
2	49.86	60.68	1.32	46
4	59.28	60.25	1.55	7
5	17.19	17.75	1.29	4
7	101.25	105.27	1.47	23
8	87.64	93.78	1.39	51
22	31.23	44.27	1.42	84

Generally, the massive use of artificial selection based on distinct breeding goals has shaped the phenotypic appearance of cattle breeds. In this context, selection signatures provide an useful information about the specific genomic regions that are conserved across generations (Mancini *et al.*, 2014; Urbinati *et al.*, 2016). The progressive breed differentiation was accelerated by introduction of breed standards, which accumulated specific phenotypes that visibly distinguished particular breeds (Rothhammer *et al.*, 2013). The analysis of evidence of recent positive selection in such breeds can provide the information on genomic regions under both artificial and natural selection and may help to identify the beneficial mutations underlying biological pathways for economically important traits (Zhao *et al.*, 2015). In comparison to our results many studies reported strong signals of positive selection in genomic regions containing QTLs for economically important traits in milk and beef cattle based on multiple approaches, including iHS (Utsunomiya *et al.* 2013; Rothhammer *et al.*, 2013; Makina *et al.*, 2015; Urbinati *et al.*, 2016).

4 CONCLUSION

The genome-wide analysis of the selective sweeps through haplotype structure of Pinzgau cattle identified several genomic regions that reflected the impact of breeding goals of Slovak and Austrian herds. The genomic regions potentially under selection showed only relatively weak signals, possibly due to the similar selection goals and close genetic connectedness between both Pinzgau populations. The results provide the first insight and basis for the future studies of genome-wide associations and impact of artificial selection on improvement of Pinzgau breeding programs.

5 ACKNOWLEDGEMENT

This study was supported by the Slovak Research and Development Agency (Contract No. APVV-14-0054 and uder contract SK-AT-2015-0016) .

6 REFERENCES

- Gautier, M., Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28, 1176–1177. Retrieved from <http://dx.doi.org/10.1093/bioinformatics/bts115>
- Gutiérrez-Gil, B., Arranz, J. J., Wiener, P. (2015). An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds. *Frontiers genetics*, 6, 167. Retrieved from <http://dx.doi.org/10.3389/fgene.2015.00167>.
- Kadlečík, O., Kasarda, R., Pavlík, I., Hazuchová, E. (2011). Pedigree Analysis of Slovak Pinzgau Breed. *Agriculturae Conspectus Scientificus*, 76(3), 165–168.
- Kadlečík, O., Swalve, H. H., Lederer, J. A. *et al.* (2004). *Development of dual-purpose Pinzgau cattle*. Nitra: SPU.
- Kasarda, R., Kadlečík, O., Ryba, Š., Bučko, O. (2009). Odhad genetických parametrov mäsovej úžitkovosti pinzgauského dobytká na Slovensku. *Acta fytotechnica et zootechnica*, 12, 37–40.
- Kasarda, R., Moravčíková, N., Trakovická, A., Mészáros, G., Kadlečík, O. (2015). Genome-wide selection signatures in Pinzgau cattle. *Potravinárstvo*, 9(1), 268–274. Retrieved from <http://dx.doi.org/10.5219/478>
- Kasarda, R., Šidlová, V., Pavlík, I., Moravčíková, N., Kadlečík, O. (2016). Inheritance of coat colour in Slovak Pinzgau cattle. *Journal of Central European Agriculture*, 17(1), 48–55. Retrieved from <http://dx.doi.org/10.5513/jcea01/17.1.1668>
- Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetic Selection Evolution*, 26, 47. Retrieved from <http://dx.doi.org/10.1186/s12711-015-0173-x>
- Mancini, G., Gargani, M., Chillemi, G., Nicolazzi, E. L., Marsan, P. A., Valentini, A., Pariset, L. (2014). Signatures of selection in five Italian cattle breeds detected by a 54K SNP panel. *Molecular Biology Reports*, 41, 957–965. Retrieved from <http://dx.doi.org/10.1007/s11033-013-2940-5>
- Mészáros, G., Kadlečík, O., Kasarda, R. (2010). Slovenský pinzgauský dobytok – dlhovekosť kráv, životnosť plemena. *Acta fytotechnica et zootechnica*, 13 (special issue), 13–15.
- Pavlík, I., Kadlečík, O., Kasarda, R., Šidlová, V., Žitný, J. (2013). Comparison of genetic diversity in dual-purpose and beef Pinzgau populations. *Acta fytotechnica et zootechnica*, 16(3), 69–73.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., Sham, P. C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81, 559–575. Retrieved from <http://dx.doi.org/10.1086/519795>
- Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G., Simianer, H. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*, 12, 318. Retrieved from <http://dx.doi.org/10.1186/1471-2164-12-318>
- Qanbari, S., Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock science*, 166, 133–143. Retrieved from <http://dx.doi.org/10.1016/j.livsci.2014.05.003>
- Rocha, D., Billerey, C., Samson, F., Boichard, D., Boussaha, M. (2014). Identification of the putative ancestral allele of bovine single-nucleotide polymorphisms. *Journal of Animal Breeding and Genetics*, 131, 483–486. Retrieved from <http://dx.doi.org/10.1111/jbg.12095>
- Rothhammer, S., Seichter, D., Förster, M., Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial

- selection in ten cattle breeds. *BMC Genomics*, 14, 908. Retrieved from <http://dx.doi.org/10.1186/1471-2164-14-908>
- 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. Retrieved from <http://dx.doi.org/10.1038/nature01140>
- Scheet, P., Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, 78, 629–644. Retrieved from <http://dx.doi.org/10.1086/502802>
- Urbinati, I., Stafuzza, N. B., Oliveira, M. T., Chud, T. C. S., Higa, R. H., de Almeida Regitano, L. C., de Alencar, M. M., Buzanskas, M. E., Munari, D. P. (2016). Selection signatures in Canchim beef cattle. *Journal of Animal Science and Biotechnology*, 7, 29. Retrieved from <http://dx.doi.org/10.1186/s40104-016-0089-5>
- Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., Van Tassell, C. P., do Carmo, A. S., Mészáros, G., Sölkner, J., Garcia, J. F. (2013). Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS One*, 8, e64280. Retrieved from <http://dx.doi.org/10.1371/journal.pone.0064280>
- Voight, B. F., Kudravalli, S., Wen, X., Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, 4, e72. Retrieved from <http://dx.doi.org/10.1371/journal.pbio.0040072>
- 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. *Genetic Selection Evolution*, 47, 49. Retrieved from <http://dx.doi.org/10.1186/s12711-015-0127-3>