

Genomic Response to Natural Selection within Alpine Cattle Breeds

NINA MORAVČÍKOVÁ^{1*}, MOJCA SIMČIČ², GÁBOR MÉSZÁROS³, JOHANN SÖLKNER³,
VERONIKA KUKUČKOVÁ¹, MICHAL VLČEK¹, ANNA TRAKOVICKÁ¹,
ONDREJ KADLEČÍK¹, RADOVAN KASARDA¹

¹Department of Animal Genetics and Breeding Biology, Slovak University of Agriculture in Nitra, Nitra, Slovakia

²Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

³Division Livestock Science, University of Natural Resources and Life Sciences Vienna, Vienna, Austria

*Corresponding author: nina.moravcikova1@gmail.com

ABSTRACT

Moravčíková N., Simčič M., Mészáros G., Sölkner J., Kukučková V., Vlček M., Trakovická A., Kadlečík O., Kasarda R. (2018): **Genomic response to natural selection within alpine cattle breeds.** Czech J. Anim. Sci., 63, 136–143.

The aim of this study was to analyse the genomic regions that have been target of natural selection with respect to identifying the loci responsible mainly for fitness traits across six alpine cattle breeds. The genome-wide scan for selection signatures was performed using genotyping data from totally 465 animals. After applying data quality control, overall 35 873 single nucleotide polymorphisms were useable for the subsequent analysis. The detection of genomic regions affected by natural selection was carried out using the approach of principal component analysis. The analysis was based on the assumption that markers extremely related to the population structure are also candidates for local adaptation potential of the population. Based on the expected false discovery rate equal to 10% up to 1138 loci were identified as outliers. The strongest signals of selection were found in genomic regions on BTA 1, 2, 3, 6, 9, 11, 13, and 22. Most genes located in the identified regions have been previously associated with immunity system as well as body growth and muscle formation that mainly reflect the pressure of both natural and artificial selection in respect to adaptation of analysed breeds to the local environmental conditions. The results also signalized that those regions represent a correlated selection response in way to maintain the fitness of analysed breeds.

Keywords: cattle; fitness; footprints of selection; genomics; local adaptation; PCAdapt

Modern cattle breeds are a result of selective breeding for many traits of economic and adaptive importance. Access to bovine genome sequence assemblies and high-density genotyping panels

provides remarkable resources to study the effects of domestication and selection on the architecture of the bovine genome. Organized breeding systems with large well characterized cattle populations

Supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic through the Slovak Research and Development Agency (Projects No. APVV-14-0054 and No. SK-AT-2015-0016) and partially supported by the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH) (Project No. SK 07/2016).

doi: 10.17221/62/2017-CJAS

also provide essential resources for the discovery of genes contributing to complex traits (Taberlet et al. 2011). Approaches that integrate quantitative trait loci (QTLs) mapping studies, genome-wide association studies (GWAS), and selection signatures have led the way to fine-map and localize functional mutations of many genes contributing to phenotypic diversity (Randhawa et al. 2016). In cattle, various GWAS have been carried out to analyse genetic differences for a variety of traits like production and reproduction and for susceptibility or resistance to certain diseases such as Johne's disease, bovine tuberculosis, mastitis or somatic cell score (Meredith et al. 2013).

Although the contribution of natural (adaptation) and artificial (domestication and subsequent directional selection) selection remains challenging to differentiate, the rapid expansion of genomic data generated from global sequencing and genotyping projects is providing greater insights into the impact of selection on genomes (Randhawa et al. 2016). It is known that the aspects of selection signatures depend on three main factors: time, age, and strength of selection events. Natural selection, as the process where organisms best adapted to the environment have higher genetic contribution to the future generations, acts in at least three ways, which are balancing, purifying, and positive selection (Martins et al. 2016). Local adaptation or positive natural selection is the driving force behind the adaptation of individuals to their environment (Fischer et al. 2014). On the other hand, artificial selection in cattle has resulted in divided breeds that are specialized for either dairy or beef production or raised as dual-purpose breeds. Such selection strategies are likely to have imposed selection pressures on the particular regions in the genome that controlled economically important traits as well as other important animal characteristics such as disease resistance or general immune competence (Zhao et al. 2015). In order to provide a list of variants that are potentially involved in the natural or artificial selection, genome scans measure the genetic differentiation between populations considering that extreme values correspond to candidate regions. Although high levels of differentiation could have various causes, the adaptation of individuals to their local environment is a prominent explanation to such patterns of differentiation for adaptive loci exceeding neutral expectations (Duforet-Frebourg et al. 2016).

Various statistical approaches have been developed to detect regions, targets of natural and artificial selection. One of the most frequently applied approach is based on the extreme values of the Wright's F_{ST} index that provides an estimate of genetic variability among populations (Weir et al. 2005). For selectively neutral loci the F_{ST} is determined by genetic drift that affects all single nucleotide polymorphisms (SNPs) across the genome in a similar way. In contrast, natural selection has locus-specific effects that can cause deviations in F_{ST} values at selected and linked loci (Akey et al. 2002). The genome scan based on F_{ST} can produce many false positives of selection signature signals due to the various biological and statistical reasons. There are important caveats with approaches related to F_{ST} because they require grouping individuals into populations, sometimes subjectively, resulting in the loss of important selection signals (Duforet-Frebourg et al. 2016). Consequently, Duforet-Frebourg et al. (2014) proposed an alternative approach to perform genome-wide scans of natural selection using principal component analysis (PCA) that uses multivariate evaluation for the population structure identification. The obtained correlations among genetic variants and each principal component provide a conceptual framework to identify the variants involved in local adaptation without a priori information on the population structure. The PCA based statistic implemented in *PCAdapt* R package provides three main advantages compared to the classic F_{ST} approach. It works on individual basis, the computation time is reduced in comparison to methods using the MCMC algorithms and candidate loci can be related to the different evolutionary events, which correspond to different principal components (PCs) (Duforet-Frebourg et al. 2016; Luu et al. 2016).

The aim of this study was to detect the genomic regions that have been target of natural selection in the respect to the identification of loci responsible for fitness traits across six alpine cattle breeds with a common historical origin.

MATERIAL AND METHODS

Genotyping data and quality control. The genome-wide scan of selection signatures was performed using genotyping data from 465 animals

of six cattle breeds with a common alpine origin (Brown Swiss – BS, Tyrol Grey – TG, Pinzgau – P, Cika – CK, Simmental – SM, and Piedmontese – PI). The final dataset was created by merging of new and previously published genotypes. The genotypes of the following cattle breeds were used: Pinzgau sires ($n = 19$) with the Slovak origin (Illumina BovineSNP50 v2 BeadChip; Illumina Inc., USA), Austrian Pinzgau ($n = 105$), Brown Swiss ($n = 103$), and Tyrol Grey ($n = 105$) sires described in Ferencakovic et al. (2013) (Illumina BovineSNP50 v1 BeadChip), Slovenian Cika sires and dams ($n = 26$) and dams (Illumina BovineSNP50 v2 BeadChip) described in Simcic et al. (2015), as well as Swiss Simmental ($n = 78$) and Italian Piedmontese ($n = 29$) breeds described in McTavish et al. (2013). To merge datasets consistently the consensus map with common SNPs across all breeds was created and overall 42 262 SNPs were retained in a reduced panel of loci. From these, all markers assigned to unmapped regions or with unknown chromosomal position according to the latest bovine genome assembly (Btau 4.6) and SNPs positioned to sex chromosomes were removed. The subsequent quality control of genotyping data was performed to exclude samples with more than 10% missing genotypes, autosomal loci with call rate lower than 90%, and minor allele frequency lower than 0.01. After applying quality control, the final dataset included 35 873 autosomal loci in total.

Identification of selection signatures. The detection of genomic regions affected by natural selection was carried out based on the approach adopted in R package *PCAdapt* (Duforet-Frebourg et al. 2016) according to Luu et al. (2016). To identify the signals of selection, the Mahalanobis distance test statistic as a multivariate method measuring the distance of the point from the mean, was used. The identification of outlier loci, the single nucleotide polymorphisms (SNPs) that are associated with positive selection, was based on the vector of z -scores obtained when regressing SNPs with the K principal components. Denoting by $z^j = (z_1^j, \dots, z_K^j)$ the vector of K z -scores between the j^{th} SNP and the first K PCs, the squared Mahalanobis distance is defined according to the Luu et al. (2016) as:

$$D_j^2 = (z^j - \bar{z})\Sigma^{-1}(z^j - \bar{z})$$

where:

\bar{z} = estimates of the z -score mean

Σ = covariance matrix of z -scores

Subsequently, the Mahalanobis distance was transformed into the P -values to perform the multiple hypothesis testing. To show a significant deviation from the null hypothesis, the quantile–quantile (Q–Q) plot was constructed. The false discovery rate approach (FDR) that provides a list of candidate markers with an expected proportion of false discoveries lower than specified value was applied to the data to determine the threshold of P -values. FDR checking was based on the q -value procedure that is adopted in the R package *qvalue* (Storey 2002) which transforms the P -values into the q -values and allows checking a specified value α of FDR and detection of candidate SNPs with q -values lower than specified α (Duforet-Frebourg et al. 2016; Luu et al. 2016). A SNP with an appropriate level of genome-wide significance was assigned to the genomic QTL location according to the Bovine Genome Database (<http://bovinegenome.org>).

RESULTS AND DISCUSSION

The PCA as an alternative method to identify the signal of selection adopted in the R package *PCAdapt* (Duforet-Frebourg et al. 2016) was applied to genotyping data firstly to analyse the population structure without a priori information of population subdivision and secondly to detect the loci associated with the local cattle adaptation potential. As expected, based on the origin of analysed cattle breeds, the first and the second PCs clearly divided the individuals into the six separate genetic clusters (Figure 1A). $K = 6$ as the optimal number of clusters confirmed also the decay of eigenvalues which decreased between $K = 5$ and $K = 7$. Figure 1B shows the proportion of variance explained by 10 PCs in decreasing order. Most of the variance was explained by the first five PCs.

A histogram of P -values (Figure 2A) confirmed that most of them followed the uniform distribution, and that the excess of small P -values indicated the presence of outliers. Figure 2A displays that the P -values were well calibrated since there was a mixture of uniform distribution and of a peaky distribution around 0, which corresponded to outlier loci. The distribution of the P -values was checked also by using the Q–Q plot (Figure 2B) that showed the expected uniform distribution of most of the P -values. The presence of outlier loci

doi: 10.17221/62/2017-CJAS

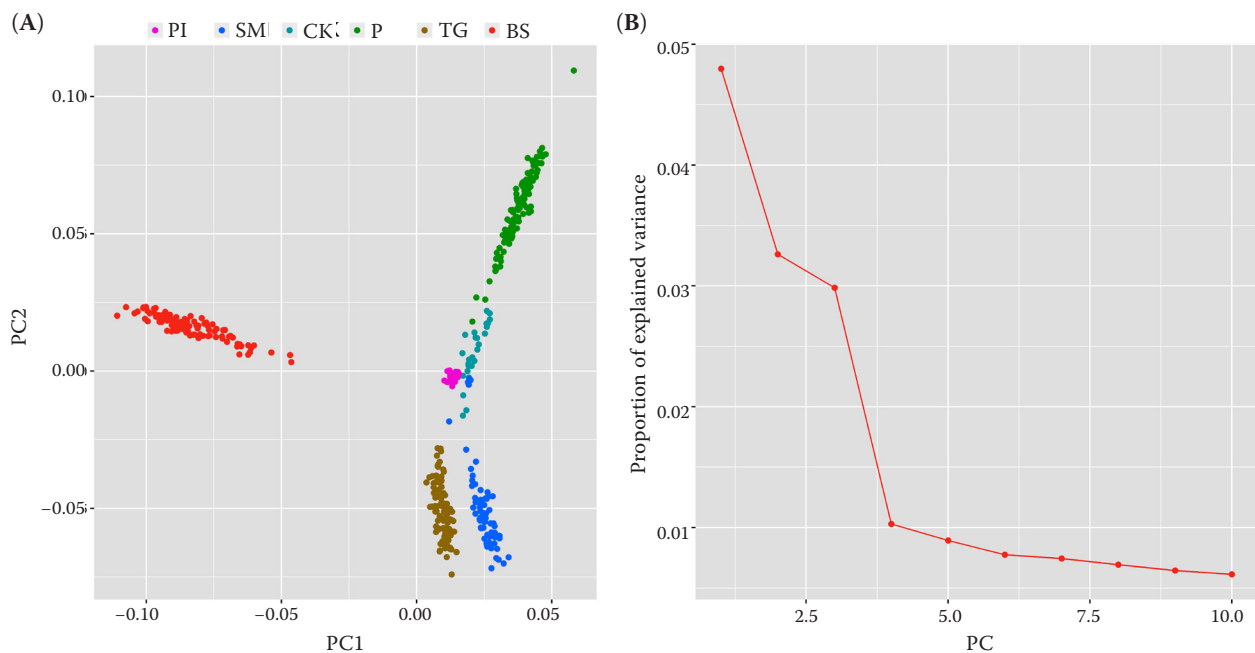


Figure 1. Scatter-plot representing the population structure across analysed breeds based on the first two principal components (PCs) (A) and the percentage of variance explained by 10 PCs (B)

PI = Piedmontese, SM = Simmental, CK = Cika, P = Pinzgau, TG = Tyrol Grey, BS = Brown Swiss

indicated the lowest P -values that were smaller than expectations.

The Manhattan plot (Figure 3) showed the main outlier loci (coloured in green) that could

be regarded as the strongest signal of selection within the six analysed cattle breeds. Based on the expected false discovery rate equal to 10% ($P = 0.01$) we were able to detect 1138 outlier

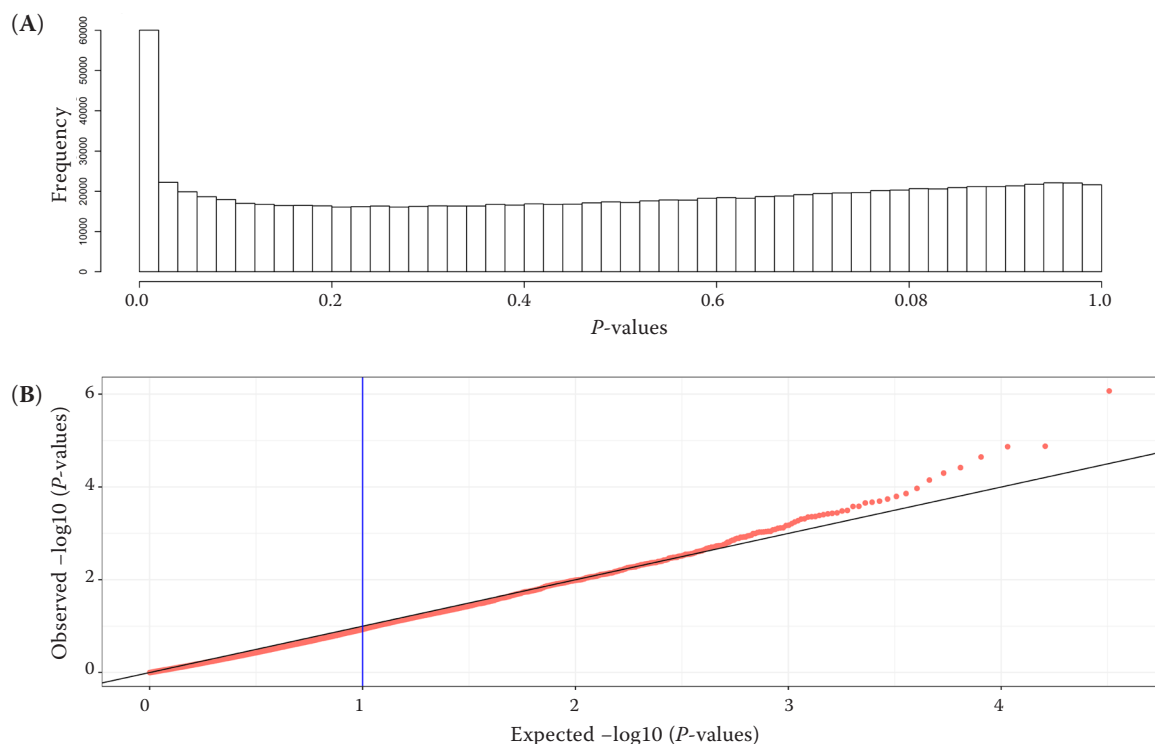


Figure 2. Distribution of P -values based on histogram (A) and the Q-Q plot (B)

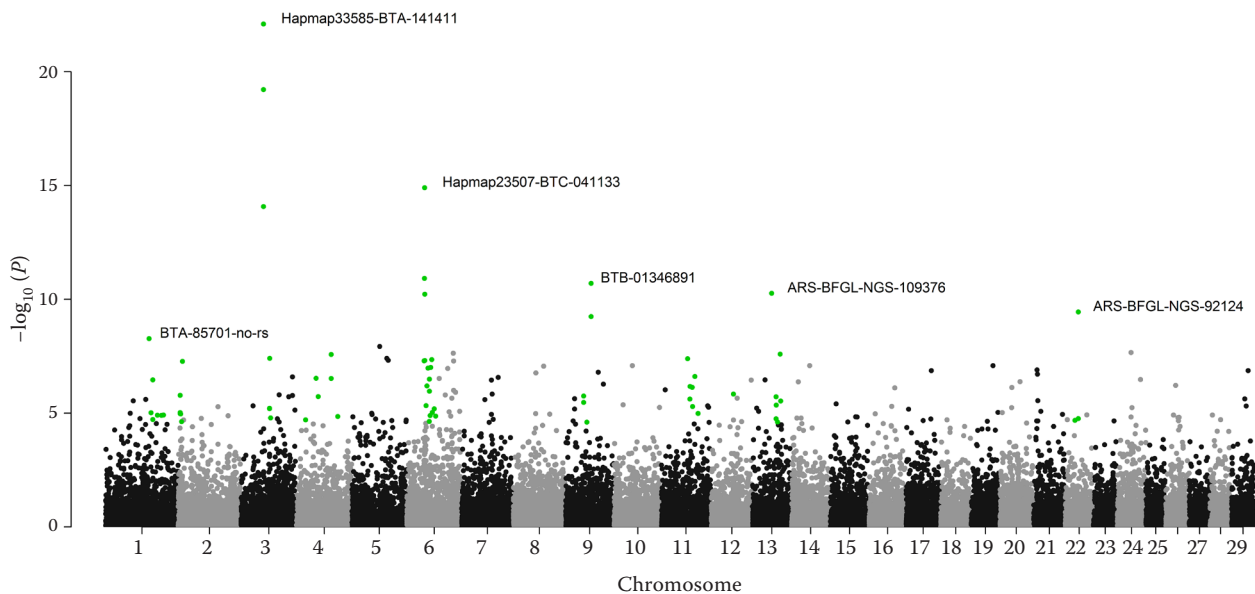


Figure 3. The Manhattan plot of $-\log_{10}(P\text{-values})$. The outlier loci characterizing the strongest signals of selection are coloured in green

SNPs distributed across all autosomes in totally 322 genomic regions. The strongest signals of selection ($P < 0.0005$) were found within eight genomic regions located on BTA 1, 2, 3, 6, 9, 11, 13, and 22 (Table 1). Most of the genes identified within these regions were previously associated mainly with immunity system (*IL12A*, *IL5RA*, *SERP1*), body growth (*GHSR*, *GFII*), and muscle formation (*MSTN*, *MYO7B*, *CAPN13*). Even if all the studied SNPs within a particular region did not reach cut-off value, the multiple signal of selection can be expected in such a region mainly due to the impact of artificial selection.

The presence of strong selective footprints across the bovine genome was tested in multiple cattle populations using various approaches mainly based on site frequency spectrum, population differentiation represented by F_{ST} statistic and haplotype length (extend of linkage disequilibrium) (Druet et al. 2013; Mancini et al. 2014; Zhao et al. 2015). The number of identified selective sweeps varied across different studies, depending on methodological approach and analysed populations. A large number of selective sweeps was presented by Stella et al. (2010) for five specialized dairy cattle breeds (215 regions), and also by Druet et

Table 1. Summary of genomic regions that showed the strongest signal of selection

Region	BTA	Start position (Mb)	End position (Mb)	No. of SNPs	QTL traits	No. of genes
1	1	94.59	133.94	11	Resistance to BSE; Birth weight; Adjusted weaning and yearling weight	244
2	2	4.06	7.49	15	Yearling weight; Kidney, pelvic and heart fat	41
3	3	43.69	64.87	13	Marbling score; Estimated kidney, pelvic and heart fat	159
4	6	37.87	62.69	21	Longissimus muscle area; Hot carcass weight; Birth weight; Yearling weight; Marbling score	95
5	9	40.12	56.41	10	Marbling score; Canonical conformation trait 2	110
6	11	57.13	72.84	14	Yield grade	92
7	13	41.95	56.69	9	Canonical conformation trait 9	226
8	22	20.18	28.71	8		23

doi: 10.17221/62/2017-CJAS

al. (2013) for 12 breeds of different production types (147 regions). A much lower proportion of selective footprints (16 regions) was found by Flori et al. (2009) for French dairy cattle breeds and Mancini et al. (2014) for Italian breeds. Despite the relatively lower proportion of selective sweeps identified for the breeds analysed in our study compared to others, the results indicated that the approach proposed by Duforet-Frebourg et al. (2016) could be a perspective alternative for the identification of selection signatures not only in human but also in cattle.

Each of the breeds analysed in our study is valuable for farmers due to the resistance to harsh conditions, durability, and longevity (Strapak et al. 2010; Meszaros et al. 2013). The results indicated that identified genomic regions showing selection signatures are important not only artificially but naturally to survive local (mountain) conditions. Two genes, *GHSR* and *SERP1*, responsible mainly for body growth, were found according to the NCBI database (<https://www.ncbi.nlm.nih.gov>) within the genomic region identified on BTA1. The growth hormone secretagogue receptor (ghrelin receptor, *GHSR*) is known to be involved in the control of growth hormone release by mediating the strong stimulatory effect of the endogenous ligand, ghrelin, on growth hormone secretion (Zhang et al. 2009) and the *SERP1* gene encodes the stress-associated endoplasmic reticulum protein 1 (Suarez-Vega et al. 2015). In the genomic region on BTA2, the myostatin (*MSTN*) gene necessary for the muscle formation in cattle was detected. The mutation in the *MSTN* gene was previously associated with the development of double muscling in several cattle breeds including Piedmontese (Luo et al. 2014). On the BTA11, a gene encoding calpain (*CAPN*) was found, that is highly associated with meat shear force and genetic variation of meat tenderness (Parra-Bracamonte et al. 2015).

In regards to the immunity response, two genes (*IL12A* and *IL5RA*) were identified in the specific regions on BTA1 and BTA22, respectively. It is generally accepted that genes of the immune system are often related to positive selection due to the continued resistance to pathogens. The crossbred animals and composite breeds are characterized by resistance to endo- and ectoparasites and, therefore, it is speculated that the identified selection signatures associated with the immune system genes reflect these selected traits (Urbinati et al. 2016).

The selection for disease resistance is much more complicated compared to the selection for production traits which can be measured directly or indirectly in every animal. Regarding the selection for disease resistance in livestock, it may not be ethical or cost efficient to challenge every animal with a pathogen to determine its level of disease resistance. From a genetic perspective, understanding the natural, innate, and acquired immune systems is crucial in developing selection programs for disease resistance. One of the tools suggested to improve inherited animal health is breeding for enhanced immune response. The mammalian innate immune system provides host defence against a variety of pathogens without requiring prior exposure, and genes modulating innate immunity have often been considered candidate loci for improving the host resistance to disease (Seabury et al. 2009). The immune system is regulated by several thousand genes (8–9% of the genome) that indicate its high genetic priority as a critical fitness trait providing survival of the species (Mallard et al. 2015). Moreover, adaptive immune response traits are heritable, and therefore it is possible to breed for improved immunity response and decreasing the occurrence of disease. In dairy cows with superior antibody-mediated (AMIR) and cell-mediated immune responses (CMIR) a lower occurrence of many diseases including mastitis has been demonstrated (Thompson-Crispi et al. 2014).

CONCLUSION

Despite the relatively small number of identified regions in respect to the fitness traits (totally 1138 outlier SNPs) in our study, the results provided encouraging evidence in support of a future genome-wide association study based on the involvement of phenotypic data that could lead to the estimation of genomic breeding values also for the immune response. Within the six alpine cattle breeds analysed, the strongest signals of selection were found in genomic regions controlling immunity system as well as body growth and muscle formation. Thus, the identified selection signatures reflected both natural and artificial selection pressure mainly in respect to adaptation of the analysed breeds to the local environmental conditions. The results also signaled that those regions represent a correlated selection response in way to maintain fitness of the analysed breeds.

Acknowledgement. The biological samples of Slovak Pinzgau bulls were provided by the Breeding Services of the Slovak Republic s.e. (PSSR).

REFERENCES

- Akey J.M., Zhang G., Zhang K., Jin L., Shriver M.D. (2002): Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, 12, 1805–1814.
- Druet T., Perez-Pardal L., Charlier C., Gautier M. (2013): Identification of large selective sweeps associated with major genes in cattle. *Animal Genetics*, 44, 758–762.
- Duforet-Frebourg N., Bazin E., Blum M.G.B. (2014): Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, 31, 2483–2495.
- Duforet-Frebourg N., Luu K., Laval G., Bazin E., Blum M.G.B. (2016): Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Molecular Biology and Evolution*, 33, 1082–1093.
- Ferencakovic M., Solkner J., Curik I. (2013): Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution*, 45: 42.
- Fischer M.C., Foll M., Heckel G., Excoffier L. (2014): Continental-scale footprint of balancing and positive selection in a small rodent (*Microtus arvalis*). *PLoS One*, 9, e112332.
- Flori L., Fritz S., Jaffrezic F., Boussaha M., Gut I., Heath S., Foulley J.L., Gautier M. (2009): The genome response to artificial selection: a case study in dairy cattle. *PLoS One*, 4, e6595.
- Luo J., Song Z., Yu S., Cui D., Wang B., Ding F., Li S., Da Y., Li N. (2014): Efficient generation of myostatin (MSTN) biallelic mutations in cattle using zinc finger nucleases. *PLoS One*, 9, e95225.
- Luu K., Bazin E., Blum M.G.B. (2016): pcadapt: an R package for performing genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.
- Mallard B.A., Emam M., Paibomesai M., Thomson-Crispi K., Wagter-Lesperance L. (2015): Genetic selection of cattle for improved immunity and health. *Japanese Journal of Veterinary Research*, 63, S37–S44.
- 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.
- Martins H., Caye K., Luu K., Blum M.G., Francois O. (2016): Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Molecular Ecology*, 25, 5029–5042.
- 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, E1398–E1406.
- Meredith B.K., Berry D.P., Kearney F., Finlay E.K., Fahey A.G., Bradley D.G., Lynn D.J. (2013): A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. *Frontiers in Genetics*, 4: 229.
- Meszaros G., Kadlecik O., Kasarda R., Solkner J. (2013): Analysis of longevity in the Slovak Pinzgau population – extension to the animal model. *Czech Journal of Animal Science*, 58, 289–295.
- Parra-Bracamonte G.M., Martinez-Gonzalez J.C., Sifuentes-Rincon A.M., Moreno-Medina V.R., Ortega-Rivas E. (2015): Meat tenderness genetic polymorphisms occurrence and distribution in five Zebu breeds in Mexico. *Electronic Journal of Biotechnology*, 18, 365–367.
- Randhawa I.A.S., Khatkar M.S., Thomson P.C., Raadsma H.W. (2016): A meta-assembly of selection signatures in cattle. *PLoS One*, 11, e0153013.
- Seabury C.M., Seabury P.M., Decker J.E., Schnabel R.D., Taylor J.F., Womack J.E. (2009): Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 107: 151.
- Simcic M., Smetko A., Solkner J., Seichter D., Gorjanc G., Kompan D., Medugorac I. (2015): Recovery of native genetic background in admixed populations using haplotypes, phenotypes, and pedigree information – using Cika cattle as a case breed. *PLoS One*, 10, e123253.
- Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010): Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*, 185, 1451–1461.
- Storey J.D. (2002): A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B*, 64, 479–498.
- Strapak P., Juhas P., Strapakova E., Halo M. (2010): Relation of the length of productive life and the body conformation traits in Slovak Simmental breed. *Archiv Tierzucht*, 53, 393–402.
- Suarez-Vega A., Gutierrez-Gil B., Klopp C., Robert-Granie C., Tosser-Klopp G., Arranz J.J. (2015): Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. *Scientific Reports*, 5, Article No. 18399.

doi: 10.17221/62/2017-CJAS

- Taberlet P., Coissac E., Pansu J., Pompanon F. (2011) Conservation genetics of cattle, sheep, and goats. *Comptes Rendus Biologies*, 334, 247–254.
- Thompson-Crispi K.A., Sargolzaei M., Ventura R., Abo-Ismael M., Miglior F., Schenkel F., Mallard B.A. (2014): A genome-wide association study of immune response traits in Canadian Holstein cattle. *BMC Genomics*, 15: 559.
- Urbinati I., Stafuzza N.B., Oliveira M.T., Chud T.C., Higa R.H., 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, 2–9.
- Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M., Hill W.G. (2005): Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, 15, 1468–1476.
- Zhang B., Chen H., Guo Y., Zhang L., Zhao M., Lan X., Zhang C., Pan C., Hu S., Wang J., Lei C. (2009): Associations of polymorphism within the GH5R gene with growth traits in Nanyang cattle. *Molecular Biology Reports*, 36, 2259–2263.
- 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.

Received: 2017–05–30

Accepted after corrections: 2018–01–15