Genomic signatures of positive selection with respect to the immunity-related genes in cattle

N. Moravčíková, R. Kasarda, V. Kukučková, A. Trakovická & O. Kadlecík

Slovak University of Agriculture in Nitra, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 94976 Nitra, Slovakia
nina.moravcikova@uniag.sk (Corresponding Author)

Summary

The aim of this study was to detect genomic regions that have been target of positive selection and to identify markers responsible for immune response across two local cattle breeds. In total of 189 animals representing the nucleus of Slovak Pinzgau and Slovak Spotted populations were included in the study. All of animals have been genotyped by using Illumina BovineSNP50 v2 Bead chip. In order to obtain more detailed view on genomic regions controlling adaptive and innate immune response only SNPs located on BTA6 (1.72-87.76 Mb), BTA21 (6.1-71.59 Mb) and BTA23 (6.98-30.50 Mb) were selected. Overall 3,250 SNPs were retained in reduced panel of markers. In order to identify genomic regions under selection pressure, the PCA approach was used based on the assumption that loci extremely related to the population structure will also candidates associated local adaptation of the population. Based on the false discovery rate equal to 10 % up to 213 loci were identified as outliers. Most of them were found on BTA6 (118) in the two genomic regions within sequence of genes encoding Toll-like receptors (TLR1, TLR6, TLR10) and Immunoglobulin J chain (RBPJ). On BTA21 the strongest signal was detected directly within sequence of immunoglobulin superfamily (ISLR, ISLR2). The lowest proportion of outlier loci was found on BTA23 (13) mostly within region controlling Major histocompatibility complex (MHC). Our results indicated that the signals of selection in genomic regions responsible for adaptive and innate immune response across both cattle breeds resulted mainly from the correlated selection response associated with maintaining their fitness. Compared to conventional breeds, the allele frequency differences can be expected in identified genomic regions. In the future, genetic selection for animals characterized by better immune response to environmental condition and diseases can provide health and productivity advantages and complement traditional health-maintenance methods.

Keywords: high-density genotyping data, local populations, Pinzgau, Slovak Spotted, selection

Introduction

Animal health and well-being have become increasingly important issues for both animal producers and consumers. Animal diseases causing morbidity and mortality significantly decrease profitability of animal production. Antibiotics that were once widely used to prevent or treat animal diseases are now administered more judiciously because of consumer fears of residual drugs in meat products and microbial resistance to commonly used antibiotics. For these reasons, new approaches or alternatives to addressing animal diseases are needed. One approach is genetic selection for animals resistant to diseases (Snowder et al., 2005). The
immune system is designed to provide protection from a wide range of pathogenic microorganisms, as well as cancers. In mammals, 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). 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 host resistance to disease in agricultural species (Seabury et al., 2010).

Slovak spotted and Slovak Pinzgau cattle whose origin is composite of autochthonous Carpathian Red (extinct) and Carpathian grey (extinct) from 17th to 18th century as well as Swiss Simmental and Austrian Pinzgau from 19th century common in Austro-Hungarian Empire belong to the main cattle breeds of national interest in Slovakia (Kasarda et al., 2015). From the beginning of exports from Austria Slovak Pinzgau have been allocated to mountains areas of Slovakia. Up to 90’s, the development of both breeds was positive from the point of population size and production. After 1990 size of both breeds has decreased significantly due to transformation processes in agriculture and exploitation of Holstein sires for crossbreeding. Today, Slovak Pinzgau cattle belongs to endangered breed (Kadlečík et al., 2013).

The aim of this study was to detect the genomic regions that have been target of positive selection in order to identify markers responsible for adaptive and innate immune response across Slovak Spotted and Slovak Pinzgau cattle.

Material and methods

In total of 189 animals representing the nucleus of Slovak Pinzgau and Slovak Spotted populations were included in the study. The sample of Slovak Pinzgau cattle consisted of breeding bulls (37), nucleus cows (35) and 80 cows. The sampled population of Slovak Spotted cattle consisted of 37 breeding bulls. All of animals have been genotyped in commercial lab by using Illumina BovineSNP50 v2 Bead chip (54,609 SNPs). The quality control of genotyping data was in first step carried out to filter out all markers assigned to unmapped regions or with unknown chromosomal position according to the latest bovine genome assembly (Btau 4.1). The subsequent SNP pruning was performed to exclude samples with more than 10% missing genotypes (2) and autosomal loci with call rate lower than 90% (392). Finally, the remaining markers on reduced SNP panel were filtered by position in order to obtain SNPs located in the genomic regions responsible for adaptive and innate immune response. Because the immunity-related genes are distributed in the bovine genome across different chromosomes, only genomic regions on BTA6 (Immunoglobulin J chain, Toll-like receptors), BTA21 (Immunoglobulin heavy chain gene cluster) and BTA23 (Major histocompatibility complex) were selected according to NCBI database. The final dataset has been composed from in total of 3,250 autosomal loci.

The detection of genomic regions affected by selection was performed based on the approach adopted in R package PCAdapt (Duforet-Frebourg et al., 2016) according to Luu et al. (2017). This approach is based on the principal component analysis (PCA). The detection of candidate markers results from the assumption that correlations between genetic variants and each principal component provide a conceptual framework to identify the variants involved in local adaptation. To identify the signals of selection the Mahalanobis distance test statistic was used. The identification of outlier loci, the SNPs that are associated with positive selection, was based on the vector of z-scores obtained when regressing SNPs with the K principal components. Subsequently, the Mahalanobis distance was transformed into the p-
values to perform the multiple hypothesis testing. The genome-wide significance thresholds for SNP p-values were determined by using the false discovery rate approach (FDR). The controlling of FDR (set to 10%) was based on the q-value procedure that is adopted in the R package qvalue (Storey, 2002). The FDR was defined as the percentage of false discoveries among the list of candidate SNPs (Luu et al., 2016; Dufre?Frebourg et al., 2016).

Results and discussion

As expected, the principal component analysis as a first step in applied approach showed clearly division of individuals in to the clusters according their genetic origin (Figure 1A). Because the decrease of eigenvalues between K=1 and K=3 was observed, the K=2 was chosen as optimal number of clusters. The distribution of p-values was checked by histogram (not shown) and Q-Q plot (Figure 1B). 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 (Figure 1B). The presence of outlier loci indicated the lowest p-values that were smaller than expectations.

Based on the false discovery rate equal to 10 % we were able to detect overall 213 outlier loci. Most of them were found on chromosome 6 (118) in the two genomic regions within sequence of Toll-like receptors (TLR1, TLR6, TLR10) and Immunoglobulin J chain (RBPJ). In cattle, Toll-like receptors are critical sensors of microbial attack and effectors of the TLR dependent innate defence mechanism, enabling the host to eliminate pathogens that otherwise would cause disease or mortality (Lakshmi et al., 2016). On the other hand, Immunoglobulin J chain regulates the structure and function of IgM polymers secreted by B cells and helps to bind immunoglobulins with secretory components (Abdel-Shafy et al., 2014).

Across chromosome 21 in total of 82 outliers under positive selection were identified. The strongest signal was found directly in genomic region of immunoglobulin superfamily (Figure 1C). On the BTA21, the main immunoglobulin heavy-chain locus is localized. The basic genetic mechanism in developing immunoglobulin diversity is similar in all jawed vertebrates. Immunoglobulins are Y-shaped hetero-tetramers consisting of two identical heavy chains (IGH) and two identical light chains, either κ or λ in mammals (IGK, IGL). Both chains are functionally divided into variable and constant domains that are combined during B-cell development (Walter et al., 2016).

The lowest proportion of outlier loci was detected on chromosome 23 (13 outliers). As shown figure 1C the strongest signal was identified within the genomic regions controlling major histocompatibility complex (MHC, BoLA-DRB3). The MHC system in cattle, known as the bovine leukocyte antigen (BoLA) is highly polymorphic and forms an important component of the immune system. The BoLA-DRB3 gene is the strongest expressed gene with the highest polymorphism level of class II locus in cattle, and influences both the magnitude and epitope specificity of antigen-specific T cell responses to infectious diseases (Takeshima et al., 2014).
Figure 1. The scatter-plot representing the population structure based on the first two principal components (A), distribution of p-values based on the Q-Q plot (B) and the Manhattan plots of $-\log_{10}$ (p-values) for BTA6, BTA21, and BTA23. The outlier loci characterizing the strongest signals of selection are coloured in green.

**Conclusion**

Our results indicated that the signals of selection in genomic regions responsible for adaptive and innate immune response across both cattle breeds resulted mainly from the correlated selection response associated with maintaining their fitness. Compared to conventional breeds, the allele frequency differences can be expected in identified genomic regions. In the future, genetic selection for animals characterized by better immune response to environmental condition and diseases can provide health and productivity advantages and complement traditional health-maintenance methods.

**Acknowledgement**

This study was supported by the Slovak Research and Development Agency (APVV-14-0054) and VEGA (1/0742/17).

**List of References**
