

Variation in Linkage Disequilibrium Patterns between Populations of Different Production Types

Veronika KUKUČKOVÁ (✉)

Nina MORAVČÍKOVÁ

Radovan KASARDA

Summary

The aim of this study was to quantify the variation in linkage disequilibrium patterns between populations of Slovak Pinzgau, Austrian Pinzgau, Simmental, Charolais and Holstein. These comparisons included differences between the genetically close populations as well as between dairy and beef breed types. Total number of genes in regions with top 0.01 and 0.1 percentile was 202. The most significant SNPs for production, reproduction and functional traits were positioned in the chromosome 7, 9, 11, 14, 20 and 24 (H2AFY, MAP3K, FAM110B, UBXN2B, CYP7A1, SDCBP, NSMAF, PRKAA1, PTGER4, MIR2361, CDH18 and C9). Genome scans confirmed the presence of selective sweeps in the genomic regions that harbour candidate genes that are known to affect productive traits in cattle such as CAST, COQ3, GJA1, ACYP2, SPTBN1, EML6, RTN4, MAP3K7, PLAG1, CHCHD7, PENK, PRLR, GHR, C6, C7, LIFR, MOCOS, GALNT1, COLEC12, CETN1, TYMS, YES1, NDC80, LPIN2, MYOM1, MYL12A, MYL12B and DLGAP1. Although phenotypic diversity is not sufficiently large to be detected, investigating the polymorphisms presented in the regions of the genome that are involved in breeding traits can be very useful in terms of genetic improvement.

Key words

BovineSNP50 Bead chip, breed uniqueness, Pinzgau cattle, selection signatures

Slovak University of Agriculture in Nitra, Faculty of Agrobiological and Food Resources, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 949 76, Nitra, Slovak Republic

✉ e-mail: veron.kukuckova@gmail.com

Received: May 16, 2017 | Accepted: August 11, 2017

ACKNOWLEDGEMENTS

This study was supported by the Slovak Research and Development Agency under the Contract No. APVV-14-0054.

Introduction

During the past 50 years the worldwide cattle population has increased approximately 50%, whereas, the production (milk, meat and hides) of this population has increased up to 100% (Randhawa et al., 2016). Genetic conditions are breed specific, given that cattle breeds were developed in relative genetic isolation and independently of each other. Until the advent of modern molecular biology methods, the technology was unavailable to identify genes associated with quantitative traits and the variants within the gene that produce differences in productivity or its expression (Kasas, Kehrlí, 2016). Pinzgau cattle kept under European conditions traditionally belong to the dual purpose breeds. Preservation of the dual purpose production type of the European Pinzgau population is generally regarded as a basic breeding aim. This breed is characterized by many excellent features, and that is why it has spread from Austria across the whole world. In spite of the fact that the Pinzgau cattle are now endangered population in Europe, many farmers are still interested in its keeping (Kasarda et al., 2016).

Signatures of selection are regions in the genome that have been preferentially increased in frequency and fixed in a population because of their functional importance in specific processes. These regions can be detected because of their lower genetic variability and specific regional linkage disequilibrium (LD) patterns (Pérez O'Brien et al., 2014). The aim of genome-wide association studies (GWAS) of common diseases and complex traits is to find statistically significant markers, usually single nucleotide polymorphisms (SNPs), which are associated with the phenotype of interest. These SNPs are seldom the genetic variants responsible for the phenotype, but are markers in LD with the underlying causal variants. Thus, the findings from GWAS serve as indicators for genomic regions that are likely to possess the genetic variants directly responsible for the functional changes and differences in phenotypic expression (Ong, Teo, 2010).

The initial discovery phase of GWAS aims to identify SNPs putatively associated with the trait of interest; the success of this phase depends on the effect size, the sample size of the studies and the significance level. Then the confirmation or validation phase is performed to assess the reproducibility of the initial association signals in the same or different populations with varying patterns of LD in different populations. Finally the fine-mapping phase is carried out to locate the functional variant which is directly responsible for phenotypic variation. In this case, the inter-population variation in LD has considerable impact on GWASs (Li et al., 2013).

The hotspots of selection could be linked to various biological functions enriched across those breeds, especially related to adaptation (disease, climate, feed resources), appearance (polledness, coat colours) and production (milk, meat, fertility) traits, each of which have economic importance in various environments and production systems (Randhawa et al., 2016).

The variation in LD method was successfully used in detection of genes associated with civilization diseases (Teo et al., 2009; Ong et al., 2010; Li et al., 2013) as well as production and reproduction traits in cattle (Perez O'Brien et al., 2014; Sorbolini et al., 2015; Randhawa et al. 2016).

The purpose of this study was to access selection signatures through linkage disequilibrium variation among Slovak Pinzgau, Austrian Pinzgau, Simmental, Holstein and Charolais. Genome signatures of production and adaptation traits was found by comparing the differences across different production types.

Material and methods

Dataset containing 381 individuals of 5 European breeds was used in further analysis. A total of 37 Slovak Pinzgau bulls (SP), 105 Austrian Pinzgau bulls (AP; Ferencaković, 2013), 84 Simmental from Switzerland (SIM), 100 Holstein from Netherlands (HOL), and 55 Charolais from France (CH) were genotyped (unknown sex; McTavish et al., 2013a). Quality control according McTavish et al. (2013b) also for Slovak and Austrian Pinzgau was performed. Subsequently, individuals with >10% missing genotypes across the autosomal markers as well as markers missing in >10% of individuals were removed. Markers with highly significant deviations ($P < 10^{-5}$) from Hardy-Weinberg Equilibrium were also excluded using the PLINK software (Purcell et al., 2007). Final number of individuals and SNP count used in the analyses was 358 and 41,135, respectively.

Population differences in local patterns of LD around SNPs associated with production and reproduction traits were measured with the varLD software (Ong, Teo, 2010), using the targeted option for 50-SNP windows, minor allele frequency (MAF = 0.05) and missingness of SNPs (0.02). Pérez O'Brien et al. (2014) concluded that unbiased estimates of LD were obtained provided $MAF > 0.05$ unless low density SNP coverage assays were used. For each population and genomic region, varLD builds a matrix of pairwise signed LD (r^2) values among all the SNP pairs and provides a raw score corresponding to the absolute difference in the eigen-decompositions between two matrices. This score is a summary measure of the overall LD levels in a given genomic region between two populations. As the magnitude of the raw varLD scores is affected by the size of the windows and the populations being compared, we prefer to use the standardized score with the empirical mean and variance of the collection of scores across the genome. This produces a statistical significance for testing the null hypothesis that the LD patterns in the region are identical between the two populations (Teo et al., 2009).

Candidate regions under positive selection by comparing genome-wide LD variation between populations were calculated for Slovak Pinzgau (SP, dual purpose), Austrian Pinzgau (AP, dual purpose), Simmental (SIM, dual purpose), Holstein (HOL, dairy) and Charolais (CHA, beef). Genomic regions containing the top 0.01, 0.1, 1 and 5 percentile of signals were characterized using the UMD3.1 *Bos taurus* genome assembly (Aken et al., 2016) to identify genes in those regions and compared with previously reported selection signatures. Data manipulation and visualisation were performed using SAS 9.3 (SAS Institute, 2011) and R software (R Core Team, 2016).

Results and discussion

The genome-wide LD variation between populations were compared to detect the candidate regions under positive selection. For all comparisons, the regions with top 0.01 and 0.1 percentile included 202 genes, including H2AFY, MAP3K,

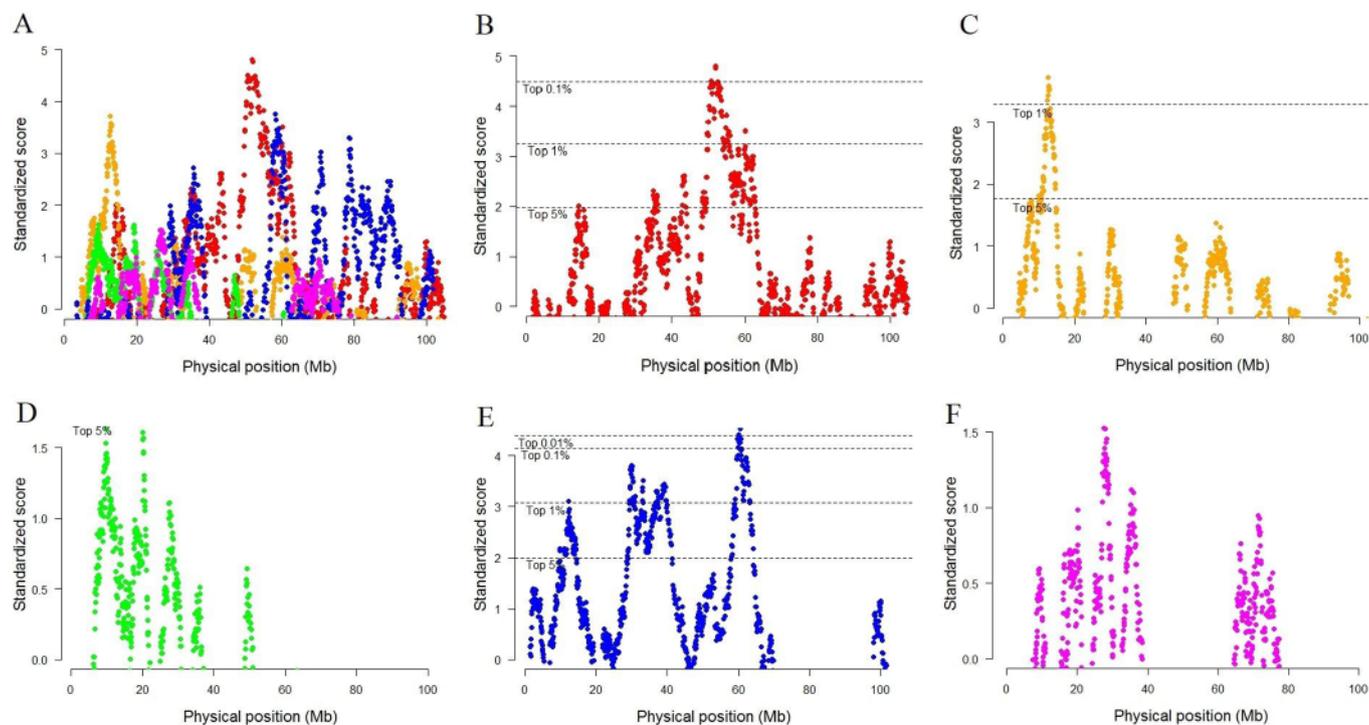


Figure 1. Normalized varLD scores in the top 5, 1, 0.1 and 0.01% of the genome-wide distributions for comparisons between pairs of populations on chromosome 11. A. The comparison of all evaluated breed pairs. B. Slovak Pinzgau – Austrian Pinzgau in red colour. C. Slovak Pinzgau – Charolais in orange colour. D. Slovak Pinzgau – Holstein in green colour. E. Slovak Pinzgau – Simmental in blue colour. F. Holstein – Charolais in violet colour.

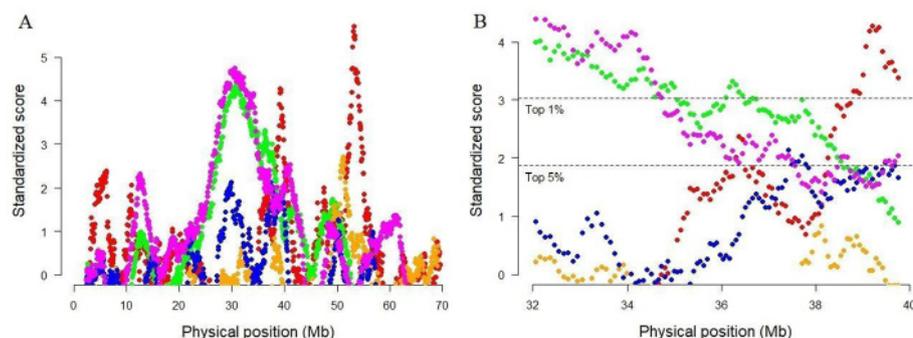


Figure 2. LD variation at the GHR (31.89-32.06 Mb) and PRLR (38.95-39.15 Mb) gene on chromosome 20. (Red points) LD comparisons between Slovak and Austrian Pinzgau; (orange points) LD comparisons between Slovak Pinzgau and Charolais; (green points) LD comparisons between Slovak Pinzgau and Holstein; (blue points) LD comparisons between Slovak Pinzgau and Simmental; (violet points) LD comparisons between Holstein and Charolais. (Dotted lines) Values of the corresponding thresholds.

FAM110B, UBXN2B, CYP7A1, SDCBP, NSMAF, PRKAA1, PTGER4, MIR2361, CDH18 and C9. As the chromosome 11 signalled a significant variation of LD in most pairs of breeds, the differences among different production types were obtained in expression of a specific region of the genome (Figure 1). In the region with the highest percentile (0.01) was found the gene associated with the feed efficiency and growth (MAP3K7) in SP and SIM pair (Figure 2E). According Sorbolini et al. (2015) has the high-quality meat big importance for human consumption. They analysed in their study selection signatures between two beef cattle breeds, Piemontese and Marchigiana. Similarly, SP as well as SIM in presented study are dual purpose breeds for

which on the level of variance in LD is possible to find the important signatures.

Randhawa et al. (2016) localized any hotspots of selection in genomic regions with medium-to-high gene density. However, some prominent peaks within low-to-medium density genic regions indicate that selection targeted a few or single genes of major effects, for example, MSTN (BTA 2), ABCG2 (BTA 6), NCAPG-LCORL (BTA 6), PLAG1-CHCHD7 (BTA 14), and GHR (BTA 20). Several studies in cattle have described and confirmed QTL with impacts on milk production, reproductive, and udder traits on bovine chromosome 20 (Viitala, 2006; Rahmatalla et al., 2011; Randhawa et al., 2016). The bovine growth hormone

Table 1. Distribution of varLD scores for the 16 autosomal candidate regions with strongest signals for natural selection in the evaluated populations (top 0.01, 0.1 and 1 percentile)

Region	Chr: start–end (Mb)	Comparison populations	Genes in region
1	Chr7: 45.53-51.68	SP – HOL	TCF3, REXO1, SHROOM1, GDF9, VDCA1, JADE2, H2AFY, MYOT, BRD8, EGRI
2	Chr7: 98.39-98.63	SP – HOL	CAST
3	Chr9: 2.72-5.1	SP – SIM	–
4	Chr9: 51.01-51.19	SP – SIM	PNISR, COQ3, FAXC
5	Chr11: 1.24	SP – SIM	KCNQ5
6	Chr11: 29.45-30.95	SP – SIM	GJA1, TBC1D32
7	Chr11: 32.95-33.29	SP – SIM	SLC35F1
8	Chr11: 36.82-39.79	SP – SIM	ACYP2, SPTBN1, EML6, RTN4
9	Chr11: 59.04-63.13	SP – SIM	MAP3K7
10	Chr14: 18.1-27.97	SP – CHA	PLAG1, CHCHD7, PENK
11	Chr20: 39-39.74	SP – AP	PRLR, AGXT2, RAD1
12	Chr20: 52.37-54.47	SP – AP	CDH18
13	Chr20: 29.41-40.51	SP – HOL	GHR, PRKAA1, PTGER4, PRLR
14	Chr20: 29.82-39.67	HOL – CHA	GHR, C6, C7, PRKAA1, PTGER4, C9, LIFR
15	Chr24: 20.31-22.16	SP – SIM	FHOD3, MOCOS, GALNT1
16	Chr24:	SP – SIM	COLEC12, CETN1, TYMS, YES1, NDC80, LPIN2, MYOM1, MYL12A,B, DLGAP1

receptor (GHR) gene has been identified as a strong positional and functional candidate gene influencing milk production. Rahmatalla et al. (2011) found also the association with milk yield, as well as with fat, protein, and casein contents. Among candidates in close neighbourhood is the prolactin receptor gene (PRLR). PRLR gene variants have shown associations with fat and protein yields in Finnish Ayrshire dairy cattle. According to the human and mouse genomic sequences, the receptor for prolactin hormone locates ~7 Mb from the GHR. Both GHR and PRLR have a major role in the regulation of growth hormone and prolactin action in the mammary gland as well as in a variety of tissues and are thus potential candidate genes that could be responsible for QTL effects observed in chromosome 20 (Viitala et al., 2006). Moreover, the interactions of the GHR gene with other genes on the same or other chromosomes can also contribute to the observed effects (Rahmatalla et al., 2011).

Figure 2A shows the normalized scores from the genome-wide assessment of LD variation in selected pairs of populations, indicating strong evidence of inter-population LD variations on chromosome 20. The genomic region starting with GTR gene and ending with PRLR gene is shown in figure 2B. Since the breed pairs with HOL are strongly peaked in GTR gene, the SP-AP pair is oppositely peaked in PRLR.

As shown in Table 1, the most significant SNPs for production, reproduction and functional traits were positioned in the chromosome 7, 9, 11, 14, 20 and 24. In the SP and SIM comparison, three genes possibly related with variation in muscle accretion were identified (MYL12A, MYL12B and MYOM1). Similar results presented also Pérez O'Brien et al. (2014) between dairy and beef cattle.

Conclusions

A total of six pair-wise comparisons between the five breeds were conducted. These comparisons included differences between the genetically close populations, between the dairy and dual-purpose breed type, between the beef and dual purpose as well as between dairy and beef breed types. The VarLD method

has the potential to capture recent strong selection because LD breaks down quickly over longer distances and, thus, high LD over an extended region is likely the result of recent selection. Genomic regions associated with traits of economic and biological importance in cattle reflected candidate genes with major effect on milk yield and composition (GHR), reproduction (PRLR), feed efficiency (PRKAA1, MOCOS), fertility (PLAG1) and immune response (COQ3, PTGER4). This study helps to explain the genetic architecture for production, reproduction and functional traits in Slovak as well as Austrian Pinzgau, Holstein, Simmental and Charolais cattle, and it provides useful information for marker-assisted selection.

References

- Aken B. L., Ayling S., Barrell D., Clarke L., Curwen V., Fairley S., Banet J. F., Billis K., Girón C. G., Hourlier T., Howe K., Kähäri A., Kokocinski F., Martin F. J., Murphy D. N., Nag R., Ruffier M., Schuster M., Tang Y. A., Vogel J.-H., White S., Zadissa A., Flicek P., Searle S. M. J. (2016). The Ensembl gene annotation system Database, baw093. doi: <http://dx.doi.org/10.1093/database/baw093>
- Casas E., Kehrli Jr. M. E. (2016). A Review of Selected Genes with Known Effects on Performance and Health of Cattle. *Front Vet Sci* 3:113. doi: <http://dx.doi.org/10.3389/fvets.2016.00113>
- Kasarda R., Moravčíková N., Šidlová V., Krupová Z., Krupa E., Kadlečík, O. (2016). Progress in evaluation of diversity in Pinzgau cattle based on molecular markers. *Archiva Zootechnica* 19(1): 37-44.
- Li H., Teo Y. Y., Tan E. K. (2013). Patterns of Linkage Disequilibrium of LRRK2 across Different Races: Implications for Genetic Association Studies. *PLoS ONE*, 8(9): e75041. doi: <http://dx.doi.org/10.1371/journal.pone.0075041>
- McTavish E. J., Decker J. E., Schnabel R. D., Taylor J. F., Hillis D. M. (2013a). New World cattle show ancestry from multiple independent domestication events. *Dryad Digital Repository*. doi: <http://dx.doi.org/10.5061/dryad.42tr0.2>
- McTavish E. J., Decker J. E., Schnabel R. D., Taylor J. F., Hillis D. M. (2013b). New World cattle show ancestry from multiple independent domestication events. *PNAS*, 110(15): E1398-E1406. doi: <http://dx.doi.org/10.1073/pnas.1303367110>

- Ong R.T.H., Teo Y.Y. (2010). varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinf app note* 26(9): 1269-1270. doi: <http://dx.doi.org/10.1093/bioinformatics/btq125>
- Pérez O'Brien A.M., Utsunomiya Y.T., Mészáros G., Bickhart D.M., Liu G.E., Van Tassel C.P., Sonstegard T.S., Da Silva M.V.B., Garcia J.F., Sölkner J. (2014). Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol* 46(1): 19. doi: <http://dx.doi.org/10.1186/1297-9686-46-19>
- 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. *Am J Hum Genet* 81: 559-575. doi: <http://dx.doi.org/10.1086/519795>
- R Core Team. (2016). R: A Language and Environment for Statistical Computing (v3.3.2). R Foundation for Statistical Computing, Vienna. <http://www.r-project.org>
- 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(4): e0153013. doi: <http://dx.doi.org/10.1371/journal.pone.0153013>
- SAS Institute Inc. (2011). SAS/STAT Software. Cary NC: SAS Institute Inc, Version 9.3.
- Sorbolini, S., Marras G., Gaspa G., Dimauro C., Cellesi M., Valentini A., Macciotta N.P.P. (2015). Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. *Genet Sel Evol* 47(1): 52. doi: <http://dx.doi.org/10.1186/s12711-015-0128-2>
- 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 Res* 19(10): 1849-60. doi: <http://dx.doi.org/10.1101/gr.092189.109>
- Viitala S., Szyda J., Blott S., Schulman N., Lidauer M., Mäki-Tanila A., Georges M., Vilkkii J. (2006). The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics* 173: 2151-2164. doi: <http://dx.doi.org/10.1534/genetics.105.046730>

acs82_19