Journal of Central European Agriculture, 2016, 17(2), p.294-302 DOI: 10.5513/JCEA01/17.2.1704

Extent of genome-wide linkage disequilibrium in Pinzgau cattle

Úroveň celogenómovej väzbovej nerovnováhy u pinzgauského dobytka

Veronika ŠIDLOVÁ[1]*, Radovan KASARDA[1], Nina MORAVČÍKOVÁ[1], Anna TRAKOVICKÁ[1], Ino CURIK[2] and Maja FERENČAKOVIĆ[2]

[1] Slovak University of Agriculture in Nitra, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic, *correspondence: veron.sidlova@gmail.com

[2] University of Zagreb, Department of Animal Sciences, Svetošimunska 25, 10 000 Zagreb, Croatia

Abstract

The aim of this study was to describe the extent of linkage disequilibrium (LD) based on genotyping data in Pinzgau cattle population. DNA samples were obtained from 19 bulls of active Pinzgau cattle population in Slovakia. Genotyping was carried out in commercial lab using an Illumina BovineSNP50 BeadChip. The genotyping array contained 54,609 single nucleotide polymorphisms (SNPs). After quality control SNPs localized on sex chromosomes were excluded. A total of 42,248 (79.97%) SNPs were found to be polymorphic. The distribution of SNPs over entire genomes of all chromosomes was not uniform. The minor allele frequencies across autosomal loci ranged from 0.266 (BTA4) to 0.280 (BTA23) with average value 0.273±0.133. Only adjacent SNPs with distance less than 5 Mb and LD (r^2) values from 0.01 to 0.99 were used in the characterization of LD extent. After filtering, the genome coordinates of 31,063 SNP markers covered all regions of autosomes on the length 2519232 kb. The spacing across genome between adjacent SNPs was in average 46.89±47.48 kb and within autosomes ranged from 43.65 kb (BTA25) to 52.59 kb (BTA5). The distribution of SNPs on different chromosomes ranged from 633 (79.13%) on BTA29 to 1123 (83.62%) on BTA14. The average LD between adjacent markers within autosomes reached values from 0.189 to 0.234 for chromosomes BTA 29 and 21, respectively. Division of r^2-values over all 29 autosomes were performed according to their physical intermarker distances and averaged within the groups for analysis of distance effect on LD. The results of our study indicated the rapid decay of LD with increasing distance between markers. Further investigation will be oriented on evaluation of effective population size based on LD data. This could allow improvement of our knowledge about genetic diversity and its use for breed preservation of original phenotype supported by the current selection schemes and breeding programmes.

Keywords: bovineSNP50 BeadChip, genetic diversity, intermarker distance

Abstrakt

Cieľom tejto štúdie bolo popísať úroveň väzbovej nerovnováhy (LD) u pinzgauského dobytka s použitím genotypových údajov. Boli získané DNA vzorky od 19 býkov aktívnej populácie pinzgauského dobytka na Slovensku, ktoré boli ďalej spracované v komerčnom laboratóriu pomocou bovinného SNP50 čipu od firmy Illumina. Molekulárno-genetické dáta obsahovali 54 609 jednonukleotidových polymorfizmov (SNPs). Pri kontrole kvality údajov boli odstránené SNPs lokalizované na pohlavných chromozómoch, pričom 79,97% z nich (42 248 SNPs) bolo polymorfných. Ich rozmiestnenie na jednotlivých chromozómoch nebolo uniformné. Frekvencia minoritnej alely autozomálnych lokusov dosiahla priemernú hodnotu 0,273 ± 0,133, v rozmedzí od 0,266 (BTA4) do 0,280 (BTA23). Pre popísanie úrovne LD boli použité len susedné SNPs s maximálnou vzdialenosťou 5 Mb a hodnotou LD (r^2) od 0,01 do 0,99. Následne bolo získaných 31 063 markérov v celkovej dĺžke 2 519 232 kb, pokrývajúce všetky autozomálne úseky genómu. Priemerná vzdialenosť medzi susednými SNPs bola 46,89 ± 47,48 kb a stúpala od 43,65 kb (BTA25) od 52,59 kb (BTA5) na jednotlivých autozómoch s rozmiestnením SNPs od 633 (79,13%) na chromozóme 29 do 1123 (83,62%) na chromozóme 14. Priemerná LD dosahovala hodnotu 0,189 – 0,234 pre chromozóm 29 a 21, v uvedenom poradí. Na základe fyzickej vzdialenosti medzi markérmi boli hodnoty r^2 všetkých 29 autozomálnych chromozómov rozdelené do skupín a priemerné hodnoty za každú skupinu použité pre analýzu vplyvu vzdialenosti na LD. Výsledky našej práce ukázali rýchly pokles LD s rastúcou vzdialenosťou medzi markérmi. Ďalšie výskumy budú zamerané na odhad efektívnej veľkosti populácie odvodenej z LD, čo môže zlepšiť naše vedomosti o genetickej diverzite a jej použití pre zachovanie plemena v jeho originálnej podobe s podporou selekčných systémov a šľachtiteľských programov.

Kľúčové slová: bovinný SNP50 čip, genetická diverzita, vzdialenosť medzi markérmi

Detailný abstrakt

Cieľom tejto štúdie bolo popísať úroveň väzbovej nerovnováhy (LD) u pinzgauského dobytka s použitím genotypových údajov. Zachovanie genetickej diverzity v rámci populácie je zvlášť dôležité pre malé, ohrozené plemená hospodárskych zvierat. Slovenské pinzgauské plemeno je kombinovaného úžitkového typu, s obľubou chované najmä v horských a podhorských oblastiach Slovenska. Napriek jeho významným vlastnostiam (dlhá životnosť, dobrá životaschopnosť a zdravie, odolnosť, efektívne využitie dostupného krmiva) sa jeho počty v posledných rokoch rapídne znížili a v súčasnosti je považované za ohrozené. Pre zachovanie genetickej diverzity a zvýšenie efektívnej veľkosti populácie je potrebný dôkladný monitoring. Celogenómové asociačné štúdie sú pripravované s cieľom identifikácie oblastí v genóme, ktoré ovplyvňujú variabilitu fenotypových vlastností. Významnú úlohu zohráva úroveň LD v rámci genómu, ktorá determinuje počet markérov nevyhnutných k identifikácii lokusov kvantitatívnych znakov. Hodnotenie LD môže byť navyše využité pre identifikáciu chromozomálnych oblastí, ktoré boli ovplyvnené selekciou, manažmentu genetických zdrojov a diverzity. S použitím genomických technológií je možné overiť súčasné znalosti a pochopenie LD u hovädzieho dobytka, jeden z parametrov genetickej diverzity súvisiaci s efektívnou veľkosťou populácie. Pre následné analýzy boli použité vzorky DNA od 19 býkov aktívnej populácie pinzgauského dobytka na Slovensku. Molekulárno-genetické dáta, obsahujúce 54 609 jednonukleotidových polymorfizmov (SNPs), boli získané v komerčnom laboratóriu za použitia bovinného čipu od firmy Illumina. Po kontrole kvality údajov boli použité SNPs lokalizované na autozómoch, pričom 79,97% z nich (42 248 SNPs) bolo polymorfných. Ich rozmiestnenie na jednotlivých chromozómoch nebolo uniformné. Frekvencia minoritnej alely autozomálnych lokusov dosiahla priemernú hodnotu 0,273 ± 0,133, v rozmedzí od 0,266 (BTA4) do 0,280 (BTA23). Pre popísanie úrovne LD boli použité len susedné SNPs s maximálnou vzdialenosťou 5 Mb a hodnotou LD (r^2) od 0,01 do 0,99, teda 31 063 markérov v celkovej dĺžke 2 519 232 kb, pokrývajúce všetky autozomálne úseky genómu. Priemerná vzdialenosť medzi susednými SNPs bola 46,89 ± 47,48 kb a stúpala od 43,65 kb (BTA25) od 52,59 kb (BTA5) na jednotlivých autozómoch s rozmiestnením SNPs od 633 (79,13%) na chromozóme 29 do 1123 (83,62%) na chromozóme 14. Priemerná LD dosahovala hodnotu 0,189 (BTA29) – 0,234 (BTA21). Na základe fyzickej vzdialenosti medzi markérmi boli r^2 rozdelené do skupín a priemerné hodnoty za každú skupinu použité pre analýzu vplyvu vzdialenosti na LD. Výsledky našej práce ukázali rýchly pokles LD s rastúcou vzdialenosťou medzi markérmi. Ďalšie výskumy budú zamerané na odhad efektívnej veľkosti populácie odvodenej z LD, čo môže zlepšiť naše vedomosti o genetickej diverzite a jej použití pre zachovanie plemena v jeho originálnej podobe s podporou selekčných systémov a šľachtiteľských programov.

Introduction

The maintenance of genetic variability within a population is important for the management of small, endangered livestock breeds. The Slovak Pinzgau is a dual purpose endangered cattle due to significant decline of the population in recent years (Kasarda et al., 2014). Thanks to its unique traits as longevity, fertility, health, grazing ability it had been bred in mountain regions of northern Slovakia (Pavlík et al., 2013). A better monitoring system is needed to increase population size without reduction of genetic diversity (Kadlečík et al., 2011).

The variance of linkage disequilibrium (LD) or the correlation of gene frequencies r is a known function, where r^2 represents the correlations between two loci. The r^2 measure range from 0 (no disequilibrium) to 1 (complete disequilibrium). The observed correlation between alleles at different loci in the genome is a determinant parameter in many applications of molecular genetics. With the wider use of genomic technologies in animal breeding and animal genetics, it is worthwhile revising and improving the current knowledge and understanding of cattle LD (Pérez O'Brien et al., 2014). Flury et al. (2010) used this method for the estimation of LD and effective population size of a local cattle breed of Switzerland. Among livestock species, this technology has been applied most successfully in cattle, because factors such as evolutionary history, genetic structure and economics make cattle particularly suitable for the application of genome assisted selection (Nicolazzi et al., 2014). Effectiveness of genomic selection and fine mapping is determined by the level of LD across the genome. Knowledge of the range of genome-wide LD, defined as a non-random association of alleles at different loci, can provide an insight into the optimal density and location of single-nucleotide polymorphisms (SNPs) for genome-wide association studies and can be a keystone for interpretation of results from QTL mapping (Bohmanova et al., 2010).

The objective of this study was to characterize linkage disequilibrium extent based on genotyping data in Slovak Pinzgau bulls.

Materials and methods

DNA samples were obtained from 19 bulls (18 from Slovakia and 1 from Austria) active in breeding programs of Pinzgau cattle in Slovakia. Genomic DNA for each of the 19 semen samples was genotyped at a commercial lab using an Illumina BovineSNP50 Genotyping v2 BeadChip. The genotyping array contained 54 609 SNPs. The quality of the SNPs was controlled by PLINK v1.07 (Purcell et al. 2007). SNPs with minor allele frequency below 5%, with more than 10% missing genotypes and SNPs with more than 95% call rate were excluded prior to analysis. Further analyses were performed including information from 42,248 autosomal SNPs.
The estimated pairwise allelic frequencies were used to calculate the squared correlation coefficient between 2 loci (r^2). Only adjacent SNPs with r^2 values from 0.01 to 0.99 were used in the characterization of LD extent. After filtering, 31,063 SNP pairs remained. The r^2 values over all 29 autosomes were averaged in 50 kb bins up to 2000 kb based on the expectation that LD for proximal SNP pairs reflects distant breeding history while LD from distal SNP pairs would reflect near history. In addition, their physical intermarker distances assuming constant recombination rate (1 Mb = 0.01 Morgan).

Results

In this study, general LD was assessed for the Slovak Pinzgau bulls, using 54,609 SNP genotypes produced from the BovineSNP50 beadchip array. The distribution of SNPs over entire genomes of all chromosomes was not uniform. A total of 42,248 (79.97%) SNPs were found to be polymorphic (MAF>0.05). The minor allele frequencies across autosomal loci ranged from 0.266 (BTA4) to 0.28 (BTA23) with average value 0.273±0.133. After filtering, 31 063 SNP pairs remained on the length 2 519 Mbp with the average adjacent marker spacing 46.89±47.48 kb ranged from 43.65 kb (BTA25) to 52.59 kb (BTA5) within autosomes (Figure 1). The distribution of SNPs on different chromosomes ranged from 633 (79.13%) on BTA29 to 1123 (83.62%) on BTA14. The average LD between adjacent markers within autosomes reached values from 0.189 to 0.234 for chromosomes BTA 29 and 21, respectively (Figure 1) with overall mean 0.212±0.218. Despite the fact that certain chromosomes had higher LD than others, there was no relationship between the length of the chromosome or the average spacing of SNPs and the extent of the LD.

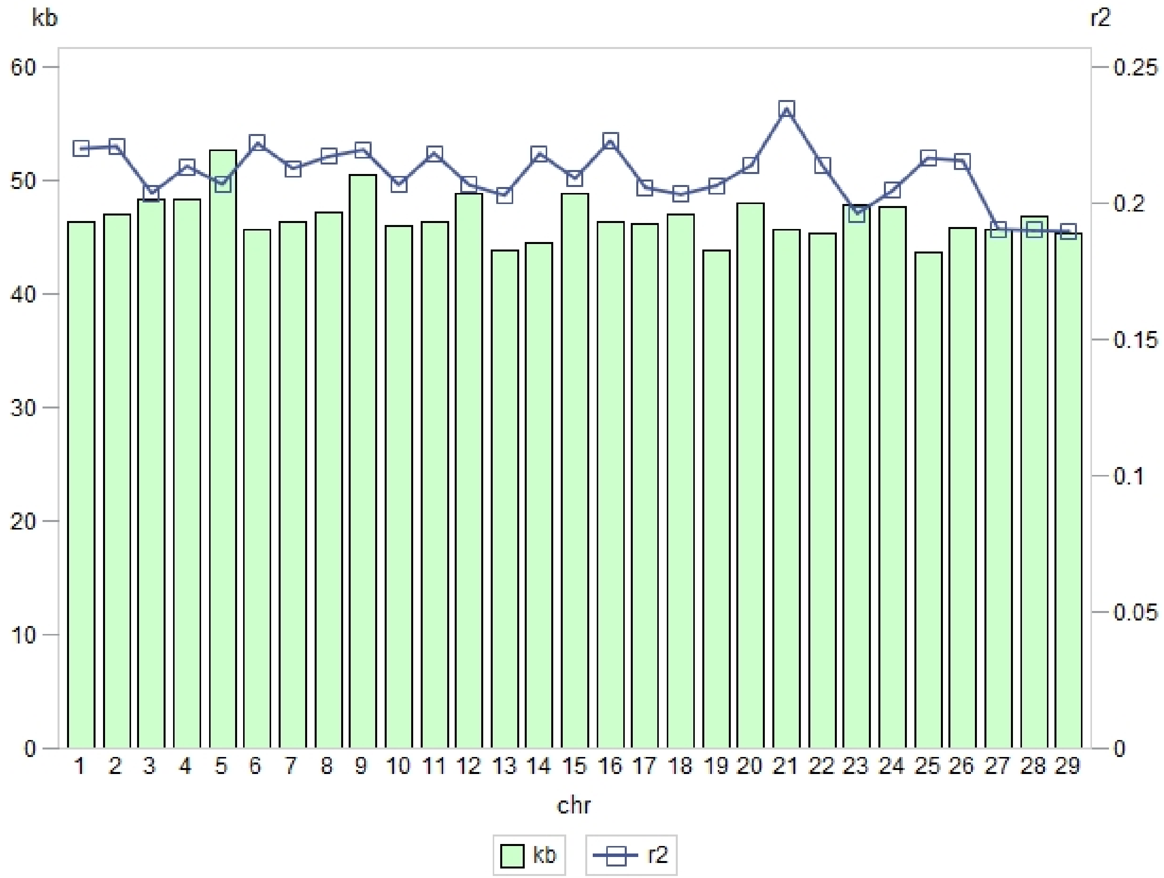

Figure 1. Distribution of average gaps and r^2-values between adjacent markers across autosomes (chr)

Obrázok 1. Rozloženie priemernej vzdialenosti a hodnoty r^2 medzi susednými markérmi na všetkých autozomálnych chromozómoch

In order to examine the decay of LD with physical distance, adjacent SNP pairs on the autosomes were sorted into bins based on their intermarker distance and average values of r^2 were calculated for each bin (Table 1). As expected most of adjacent marker pairs were included in the smallest distance bin, therefore had quite short intermarker distances. For marker pairs 0–50 kb apart, BTA21 builds the upper end with an average r^2 of 0.203 and BTA15 the lower end with the average r^2-value of 0.158. For marker pairs more than 250 kb apart, r^2 ranges from 0.025 (BTA27) to 0.42 (BTA21). This suggests that physical distance on the chromosome is not a perfect indicator of the strength of linkage. Variation in average LD between chromosomes is large for smaller distance and levels out around 200 kb. The average r^2-values among chromosomes for the first four distance bins (0-50, 50-100, 100-150, 150-200) are presented in Figure 2.

Table 1. Pairwise linkage disequilibrium (r^2) for adjacent SNPs at various distances pooled over all autosomes

Tabuľka 1. Priemerná párová väzbová nerovnováha (r^2) susedných SNPs pre rôzne vzdialenosti medzi markérmi

Bin (kb)	Pairs (n)	r^2 ± SD	Minimum	Maximum	Lower and upper 95% CI
0-50	22,618	0.225 ± 0.226	0.01	0.93	0.222 ± 0.228
50-100	6,564	0.184 ± 0.196	0.01	0.92	0.179 ± 0.188
100-150	1,226	0.155 ± 0.167	0.01	0.9	0.146 ± 0.165
150-200	378	0.144 ± 0.163	0.01	0.86	0.127 ± 0.160
200-250	154	0.138 ± 0.153	0.01	0.76	0.114 ± 0.162
< 250	123	0.117 ± 0.122	0.01	0.59	0.095 ± 0.139

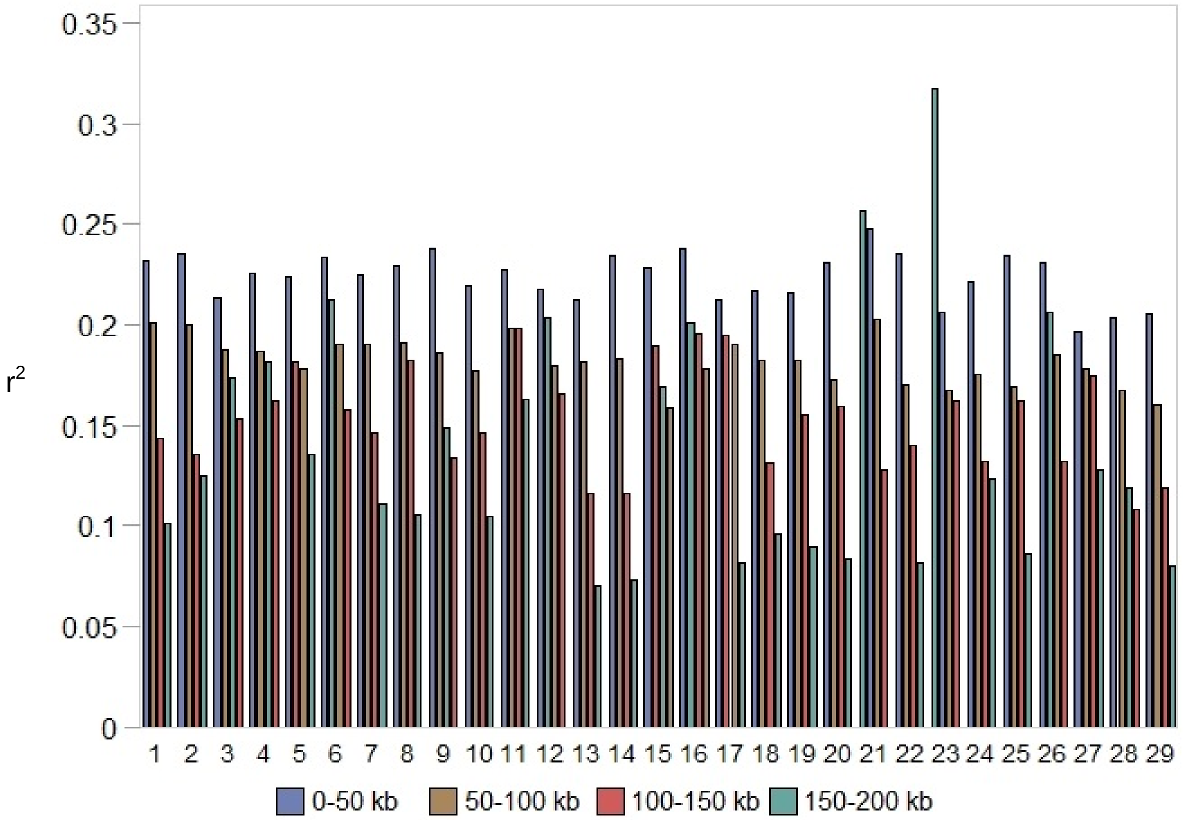

Figure 2. Chromosome-wise average r^2-values between adjacent SNPs and the decay of linkage disequilibrium with increasing physical distances for the four smallest distance bins evaluated

Obrázok 2. Priemerná hodnota r^2 medzi susednými SNPs a rozloženie väzbovej nerovnováhy s narastajúcou fyzickou vzdialenosťou v štyroch najmenších skupinách vzdialeností pre všetky chromozómy

Discussion

In this study, LD were examined using 31,063 SNP pairs genotypes in 19 individuals of the Pinzgau cattle, an indigenous cattle breed from the mountain and submountain areas of Slovakia. The average spacing across genome between adjacent SNPs was in range 43.65-52.59 kb on autosomes 25 and 5, respectively, in agreement with Uimari and Tapio (2011). Six chromosomes showed gaps between two adjacent SNPs above 1 Mb, and the largest gap was observed for BTA4 (2,940 kb) closely followed by BTA20 with 2,815 kb. In comparison with other studies assessing the LD in cattle (Khatkar et al. 2008; Sargolzaei et al. 2008), the marker density in our study is slightly higher with an average distance between adjacent markers of 46.89 kb despite studies of Flury et al. (2010) and Beghain et al. (2012) with noticeably higher average gap (75.23; 68.94). Considering the levels of LD between adjacent markers, the average r^2 in the Pinzgau (0.212) was lower than in Swiss Eringer (0.24; Flury et al., 2010) and Australian Holstein (0.27) used for genomic selection (Hayes et al. 2009), but slightly higher than in North American Holstein (0.2; Bohmanova et al., 2010) and in Blonde d´Aquitaine (0.205; Beghain et al., 2012). There is variation in the published LD extent because the estimates of LD strongly depends on various factors such as history and structure of the studied population, sample size, marker type, density and distribution of markers, type of method used, strictness of SNP filtering (Bohmanova et al., 2010).

Recent studies have shown that the level of LD in cattle is lower than previously thought and mean LD rapidly decays with increasing intermarker distance (Bohmanova et al., 2010; Flury et al., 2010; Khatkar et al. 2008; Sargolzaei et al. 2008). The results of our study support the rapid decay of LD with increasing distance between markers even using only adjacent marker pairs. Noticeable variance in LD is observed for smaller distance and levels out around 200 kb. For marker pairs 150-200 kb apart, r^2 ranges from 0.07 (BTA13) to 0.317 (BTA23). The high extent of LD across chromosomes is caused by smaller number of SNPs and few pairs located on the same genes (e.g. KHDRBS2 on chromosome 23 and LPHN3 on chromosome 6). This suggests that physical distance is not a perfect indicator of the strength of linkage, especially in the case if adjacent pairs are located on the same gene. Kim and Kirkpatrick (2009) reported a higher LD and a higher variation in intragenic regions (within genes) than in intergenic whilst Bohmanova et al. (2010) observed no differences in the extent of LD and the decline of LD with the distance between intragenic and intergenic regions.

Conclusions

The application of bovine genotyping arrays is usable tool for genetic diversity parameters estimation and presented methodology of LD extent examination is applicable for every livestock population. Further investigation will be oriented on evaluation of diversity parameters of Pinzgau and other cattle populations. This could allow improve our knowledge of genetic variability with using for preservation of the breed in the original phenotype supported by the current selection schemes and breeding programmes.

Acknowledgements

This study was supported by the Slovak Research and Development Agency under the Contract No. APVV-0636-11 and APVV-14-0054.

References

Beghain, J., Boitard, S., Weiss, B., Boussaha, M., Gut, I., Rocha, D. (2012) Genome wide linkage disequilibrium in the Blonde d'Aquitaine cattle breed. Journal of Animal Breeding and Genetics, 130 (4), 294-302. DOI: 10.1111/j.1439-0388.2012.01020.x

Bohmanova, J., Sargolzaei, M., Schenkel, F. S. (2010) Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics, 11, 421. DOI: 10.1186/1471-2164-11-421

Flury, C., Tapio, M., Sonstegard, C., Drogemuller, C., Leeb, T., Simianer, H., Hanotte, O., Rieder, S. (2010) Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. Journal of Animal Breeding and Genetics, 127 (5), 339–347. DOI: 10.1111/j.1439-0388.2010.00862.x

Hayes, B. J., Bowman, P. J., Chamberlain, A.J., Goddard, M. E. (2009) Genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science, 92 (2), 433-443. DOI: 10.3168/jds.2008-1646

Kadlečík, O., Kasarda, R., Pavlík, I., Hazuchová, E. (2011) Pedigree Analysis of Slovak Pinzgau Breed. Agriculturae Conspectus Scientificus, 76 (3), 165-168.

Kasarda, R., Mészáros, G., Kadlečík, O., Hazuchová, E., Šidlová, V., Pavlík, I. (2014) Influence of mating systems and selection intensity on the extent of inbreeding and genetic gain in the Slovak Pinzgau cattle. Czech Journal of Animal Science, 59 (5), 219-226.

Khatkar, M. S., Nicholas, F. W., Collins, A. R., Zenger, K. R., Cavanagh, J. A., Berris, W., Schnabel, R. D., Taylor, J. F. Raadsma, H. W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics, 9, 187. DOI: 10.1186/1471-2164-9-187

Kim, E. S., Kirkpatrick, B. W. (2009) Linkage disequilibrium in the North American Holstein population. Animal Genetics, 40 (3), 279-288. DOI: 10.1111/j.1365-2052.2008.01831.x

Nicolazzi, E., Picciolini, M., Strozzi, F., Schnabel, R. D., Lawley, C., Pirani, A., Brew, F., Stella, A. (2014) SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics, 15, 123. DOI: 10.1186/1471-2164-15-123

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.

Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, J. F., Van Tassell, C. P., Carvalheiro, R., da Silva, M. V. B., Solkner, J. (2014) Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. Livestock Science, 166, 121-132. DOI: 10.1016/j.livsci.2014.05.007

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., Sham, P. C. (2007) PLINK: a tool set for whole genome association and population-based linkage analysis. The American Journal of Human Genetics, 81 (3), 559-575. DOI: 10.1086/519795

Sargolzaei, M., Schenkel, F. S., Jansen, G. B., Schaeffer, L. R. (2008) Extent of linkage disequilibrium in Holstein cattle in North America. Journal of Dairy Science, 91 (5), 2106-17. DOI: 10.3168/jds.2007-0553

Uimari, P., Tapio, M. (2011) Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. Journal of Animal Science, 89 (3), 609-614. DOI: 10.2527/jas.2010-3249