

# GENETIC DIFFERENTIATION OF SLOVAK PINZGAU, SIMMENTAL, CHAROLAIS AND HOLSTEIN CATTLE BASED ON THE LINKAGE DISEQUILIBRIUM, PERSISTENCE OF PHASE AND EFFECTIVE POPULATION SIZE

Veronika KUKUČKOVÁ <sup>1</sup>, Nina MORAVČÍKOVÁ <sup>2</sup>, Anna TRAKOVICKÁ <sup>3</sup>, Ondrej KADLEČÍK <sup>4</sup>, Radovan KASARDA <sup>5</sup>

## ABSTRACT

The aim of this study was to estimate the linkage disequilibrium extent (LD), effective population size ( $N_e$ ) and the persistence of phase ( $R$ ) among cattle breeds for syntenic marker pairs using genome-wide single-nucleotide polymorphisms genotyped in Slovak Pinzgau, Simmental, Charolais and Holstein cattle. The average LD ranged from  $0.29 \pm 0.23$  (Charolais) to  $0.36 \pm 0.29$  (Simmental) for markers in the 0 to 10 kb bins. The recent  $N_e$  was from 58 (Holstein) to 85 (Charolais). The persistence of phase as a measure of the prediction reliability for markers in one breed compare to another breeds markers was estimated. Markers in the 0 to 10 kb bin could be predicted with a maximal accuracy of 0.65 between Simmental and Charolais. The lowest correlation was observed between Slovak Pinzgau and Charolais (0.46 in the 0 to 10 kb bins) what indicated that phase was not strongly preserved between breeds. Results of this study could be the basis for the implementation of genomic selection programs in the Slovak Pinzgau breed and the differentiation of cattle populations during the selection process.

**Key words:** endangered breed, selection criteria, BovineSNP50 BeadChip, linkage disequilibrium, effective population size, persistence of phase

## 1 INTRODUCTION

Knowledge about the extent and the pattern of linkage disequilibrium (LD) in livestock populations is essential to determine the density of single nucleotide polymorphisms (SNPs) required for accurate genome-wide association and genomic selection studies (Biegelmeyer *et al.*, 2016). 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 genetics, it is worthwhile revising and improving the current knowledge and understanding of cattle LD (Pérez O'Brien *et al.*, 2014). LD is common between alleles at neighbouring loci that tend to be inherited together and associated in a segregating population (Du

*et al.*, 2007). Moreover, observed LD is related to historical effective population sizes ( $N_e$ ), and can provide insights into the genetic diversity history of populations (Biegelmeyer *et al.*, 2016). The square correlation of alleles at two loci ( $r^2$ ) is considered the most robust measure of LD (Badke *et al.*, 2012).

Persistence of phase ( $R$ ) is a measure of the degree of agreement of LD phase for pairs of SNP between two populations (Badke *et al.*, 2012). Persistence of allele phase between breeds can be used to identify the history and the relationship among breeds within a species or to determine the marker density to conduct multi-breed genomic selection (Makina *et al.*, 2016).

If  $r^2$  between two markers is equal in two populations, but their corresponding  $r$  has the opposite sign, the gametic

<sup>1</sup> Slovak University of Agriculture in Nitra, Faculty of Agrobiolgy and Food Resources, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 949 76, Nitra, Slovak Republic, e-mail: veron.sidlova@gmail.com

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

<sup>3</sup> Same address as 1, e-mail: anna.trakovicka@uniag.sk

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

<sup>5</sup> Same address as 1, e-mail: radovan.kasarda@uniag.sk

phase is reversed. High positive values are a result of high  $r^2$  and equal phase in both breeds (Uimari and Tapio, 2011).

The Slovak Pinzgau breed is a dual purpose endangered breed due to a significant decline of the population in the recent years (Kasarda *et al.*, 2014). Thanks to its unique traits as longevity, fertility, health and grazing ability it is bred in the mountain regions of northern Slovakia. The population faced the bottleneck effect and the loss of diversity due to unequal use of founders. Further population size decreasing can lead to serious problems. Therefore, it is necessary to preserve the recent population in both types (dairy and beef; Pavlík *et al.*, 2013). A better monitoring system is needed to increase the population size without a genetic diversity reduction (Kadlečík *et al.* 2011).

The objective of this study was to estimate LD levels, persistence of phase and effective population size in Slovak Pinzgau breed in comparison with Simmental, Charolais and Holstein cattle.

## 2 MATERIAL AND METHODS

All Pinzgau sires ( $n = 19$ ) in Slovakia (PIN) used in insemination were analysed. Genomic DNA was isolated from the semen samples and genotyped at a commercial lab using the Illumina BovineSNP50v2 BeadChip (Illumina Inc., San Diego, CA). The genotyping array contained 54,609 SNPs. Genotyping information (BovineSNP50v2 BeadChip) for Simmental (SIM;  $n = 84$ ), Charolais (CHA;  $n = 55$ ) and Holstein (HOL;  $n = 100$ ) breeds were retrieved from the public digital repository (McTavish *et al.*, 2013b). For the retrieved dataset the quality control has already been performed according to McTavish *et al.* (2013a). The same criteria were used in Slovak Pinzgau genotyping data. Mark-

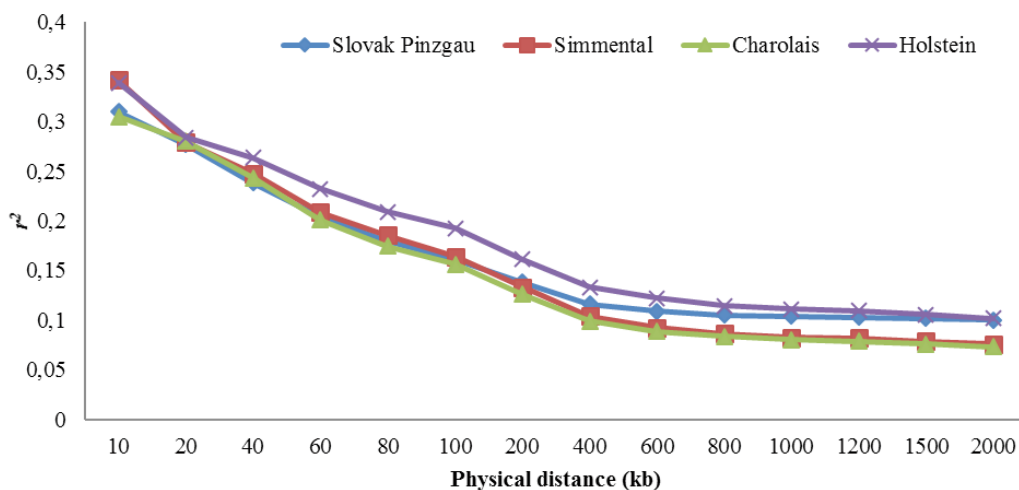
ers and individuals with call rates less than 90 % were excluded, as well as markers with highly significant deviations ( $p < 10^{-5}$ ) from Hardy-Weinberg Equilibrium were also removed using PLINK v1.9 (Purcell *et al.*, 2007).

Pérez O'Brien *et al.* (2014) observed no evidence of a minor allele frequency (MAF) influence on  $r^2$  estimates and concluded that unbiased estimates of LD were obtained provided  $MAF > 0.05$  unless low density SNP coverage assays were used. Consequently, the MAF threshold in this study was set to 0.05.

Genome-wide LD as  $r^2$  for different genetic distances between SNPs were calculated for PIN (dual purpose and dairy breed), SIM (dual purpose breed), CHA (beef breed) and HOL (dairy breed). The persistence phase of LD (R) between alleles on the same chromosome was calculated as the Pearson correlation coefficient for marker pairs that were in common within four breeds. The correlations among  $r$  values (the root of LD estimator  $r^2$ ) between breeds were estimated for 10 kb intervals (from 0 to 500 kb) using PROC CORR command in SAS 9.3 programme (SAS Institute, 2011).

## 3 RESULTS AND DISCUSSION

Genome-wide LD as  $r^2$  for different distances between SNPs was calculated for PIN, SIM, CHA and HOL. Average LD for all breeds is presented in Figure 1. The average  $r^2$  was from  $0.29 \pm 0.23$  (CHA) to  $0.36 \pm 0.29$  (SIM) and the average  $r$  was from  $0.49 \pm 0.23$  (CHA) to  $0.54 \pm 0.26$  (SIM) for markers in 0 to 10 kb bin. All studied breeds showed an inverse relationship between LD and inter-marker distances. The decline of LD as a distance function agreed with other studies based on the  $r^2$  estimates in cattle (Du *et al.*, 2007;



**Figure 1:** Decrease of linkage disequilibrium ( $r^2$ ) by physical distance in studied breeds

**Table 1:** Correlations of allele phase among all evaluated pairs of breeds by physical distance

Breed pairs	bins distances (kb)					
	10	20	50	100	200	500
Slovak Pinzgau - Simmental	0.5899	0.4415	0.4315	0.3582	0.1885	0.0666
Slovak Pinzgau - Charolais	0.4561	0.4020	0.4483	0.3541	0.1868	0.0585
Slovak Pinzgau - Holstein	0.5687	0.4318	0.3574	0.2860	0.1676	0.0111
Simmental - Charolais	0.6497	0.5524	0.5486	0.4871	0.3088	0.0904
Simmental - Holstein	0.6018	0.4184	0.3842	0.3224	0.2521	0.0268
Charolais - Holstein	0.5191	0.4570	0.4460	0.3373	0.2501	0.0427

Badke *et al.*, 2012; Pérez O'Brien *et al.* 2014; Garcia-Ruiz *et al.*, 2015; Biegelmeyer *et al.*, 2016; Makina *et al.*, 2016) where was shown that LD is a population specific. Therefore, it could be heterogeneous among populations depending on the demographic history of the populations.

Past and recent effective population sizes were estimated from the average  $r^2$  for markers separated by various genomic distances. The extent of LD over distances approximately 6000 kb indicated more recent  $N_e$  ( $N_e \approx 10$  generations ago) while that over shorter distances (60 kb) provided historical  $N_e$  ( $N_e \approx 1000$ ). Historical  $N_e$  were from 1713 (HOL) to 2304 (PIN), while more recent  $N_e$  were only from 58 (HOL) to 85 (CHA). The most rapid decline of  $N_e$  was found in PIN in the last five to six generations, confirming PIN considered as endangered breed from the year 1993.

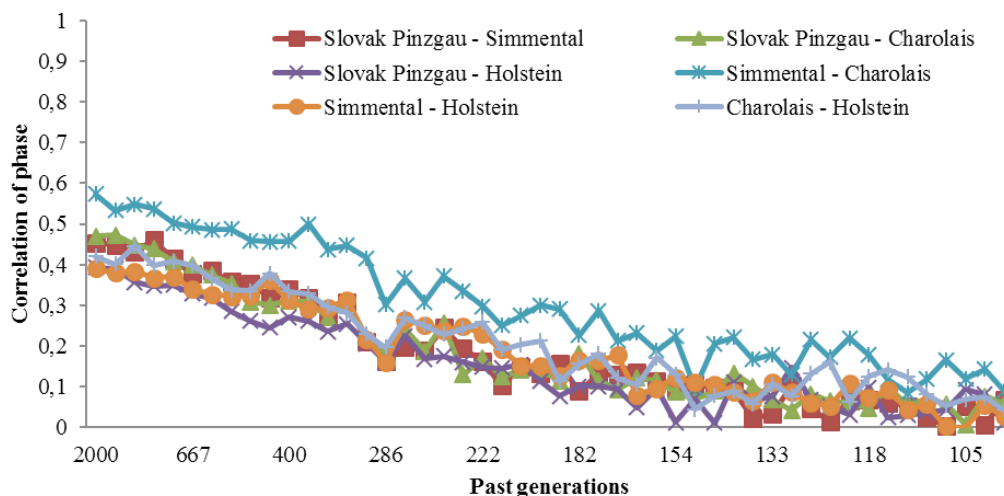
The persistence of phase between PIN, SIM, CHA and HOL populations was calculated at the interval distances of 10 kb up to 500 kb. The largest R was observed between SIM and CHA pair (0.65 in 0 to 10 kb bin), while the lowest R was observed between PIN and CHA (0.46 in 0 to 10 kb bin). However, for the distance of bins 50 kb and more, the

lowest R between PIN and HOL was obtained. The correlations between alleles at adjacent loci indicated that phase was not strongly preserved within the breed pairs and it decreased with the increasing of physical distance among breeds (Table 1).

Correlation of phase as well as  $N_e$  declined among breeds from about 2000 generations to 100 generations ago (Fig. 2). As reported Garcia-Ruiz *et al.* (2015) the persistence of LD phase was decreasing when the distance between markers was increasing. The value R refers to the correlation of LD of SNP pairs in one breed with LD of the same pairs in the other breed. The persistence of phase (correlation of  $r$  values) among populations showed the genetic relationship among them (Badke *et al.*, 2012) and thus it was used as a measure among the breeds in this study.

Observed correlations of allele phases among breeds in this study were much lower than previously reported between European cattle breeds at a short range (<10 kb; Biegelmeyer *et al.*, 2016), but higher than among South African breeds (Makina *et al.*, 2016).

The difference between the persistence of phase among


**Figure 2:** Correlations of allele phases between all evaluated breeds pairs by physical distance represented as generations in the past

the breeds in this study at the shortest distances ranged from 0.02 to 0.19, what was much higher than among Holstein at the shortest distance (0.01–0.04; Garcia-Ruiz *et al.*, 2015). Similar results were obtained among Angus, Holstein and South African breeds (Makina *et al.*, 2016) or different specie like pigs (Badke *et al.*, 2012).

Correlation between alleles at adjacent loci indicated that phase was not strongly preserved among breeds, whereas R reached moderately positive values in all population pairs in the distance < 500 kb. The statistically significant slightly negative values of R between PIN and HOL in 720 to 730 kb bin as well as between PIN and CHA in 740 to 750 kb bin were observed. According to Uimari and Tapio (2012) the only high positive values are the result of a strong  $r^2$  and an equal phase in both breeds, while negative values reveals the reversed gametic phase.

#### 4 CONCLUSIONS

Previous studies have shown that LD is a population specific and could be heterogeneous among populations depending on the demographic history of the populations. In this study, average LD was from 0.29 (Charolais) to 0.36 (Simmental) for markers in 0 to 10 kb bin. The correlation of allele phase referred to the correlation of LD values of SNP pairs in one breed with LD values of the same pairs in the other breed. The largest persistence of phase (R) was observed between Simmental and Charolais (0.65 in 0 to 10 kb bin), while the lowest one between Slovak Pinzgau and Charolais (0.46 in 0 to 10 kb bin).

The results obtained in this study, provide important details, which should be considered for the future genetic research on the Slovak Pinzgau cattle.

#### 5 ACKNOWLEDGEMENT

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

#### 6 REFERENCES

- Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C., Steibel, J. P. (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics*, 13(1), 24. doi: <http://dx.doi.org/10.1186/1471-2164-13-24>
- Biegelmeyer, P., Gulias-Gomes, C. C., Caetano, A. R., Steibel, J. P., Cardoso, F. F. (2016). Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. *BMC Genetics*, 17(32). doi: <http://dx.doi.org/10.1186/s12863-016-0339-8>
- Du, F. X., Clutter, A. C., Lohuis, M. M. (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences*, 3(3), 166–178. doi: <http://dx.doi.org/10.7150/ijbs.3.166>
- Garcia-Ruiz, A., Ruiz-López, F. de J., Van Tassel, C. P., Montaldo, H. H., Huson, H. J. (2015). Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins. *Frontiers in Genetics*, 6(7). doi: <http://dx.doi.org/10.3389/fgene.2015.00007>
- Kadlečík, O., Kasarda, R., Pavlík, I., Hazuchová, E. (2011). Pedigree analysis of Slovak Pinzgau breed. *Agriculturae Conspetus 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.
- Makina, S. O., Taylor, J. F., van Marle-Köster, E., Muchadeyi, F. C., Makgahlela, M. L., MacNeil, M. D., Maiwashe, A. (2016). Extent of linkage disequilibrium and effective population size in four South African Sanga cattle breeds. *Frontiers in Genetics*, 6(337). doi: <http://dx.doi.org/10.3389/fgene.2015.00337>
- 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. *PNAS*, 110(15), E1398–E1406. doi: <http://dx.doi.org/10.1073/pnas.1303367110>
- 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. *Dryad Digital Repository*. doi: <http://dx.doi.org/10.5061/dryad.42tr0.2>
- 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'Brian, A. M., Mészáros, G., Utsunoomiya, Y. T., Sonstegard, T. S., Garcia, J. F., Van Tassel, 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: <http://dx.doi.org/10.1016/j.livsci.2014.05.007>
- 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. doi: <http://dx.doi.org/10.1086/519795>
- SAS Institute Inc. (2011). *SAS/STAT Software*. Cary NC: SAS Institute Inc, Version 9.3.
- 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, 609–614. doi: <http://dx.doi.org/10.2527/jas.2010-3249>