



GENETIC DIVERGENCE OF CATTLE POPULATIONS BASED ON GENOMIC INFORMATION*

R. Kasarda¹, N. Moravčíková¹, G. Mészáros², A. Trakovická¹, O. Kadlečík¹

¹*Slovak University of Agriculture, Faculty of Agrobiological and Food Resources, Nitra, Slovak Republic*

²*University of Natural Resources and Life Sciences, Division of Livestock Sciences, Vienna, Austria*

The aim of this study was to analyze the extent of linkage disequilibrium (LD) and estimate the effective population size in three divergent cattle breeds. In total 19 Pinzgau, 30 Brown Swiss, and 30 Nelore bulls were genotyped using Illumina BovineSNP50 v2 BeadChip. The level of LD was assessed only for syntenic adjacent autosomal loci. The extent of LD significantly varied across the analyzed groups and autosomes. The highest average value of r^2 , which was used as measure of LD extent, was found in Brown Swiss (0.27). The r^2 values were averaged into 50 kb distance bins to estimate their relationship to SNP physical positions. With the increasing genetic distance a decay of LD was found by each of the analyzed breeds. Moreover, the decrease in relation to the distance was observed also in effective population size (N_e) estimation. Thus, the long distances in genome produced a signal of the recent history of population. The observed values of recent N_e across evaluated cattle breeds were above the minimum number required to maintain genetic diversity and indicated a sufficient animal basis for the future management of breeding programs for each of the populations.

BovineSNP50 BeadChip, effective population size, linkage disequilibrium



doi: 10.1515/sab-2016-0016

Received for publication on July 28, 2015

Accepted for publication on February 3, 2016

INTRODUCTION

The current availability of high-density SNP arrays enables the application of genomic techniques to livestock. The genome-wide association analysis in connection to the genomic selection is one of the most important genomic techniques which depend on the extent of linkage disequilibrium (LD) and its rate of decline with genetic distance between loci within a population (Shin et al., 2013). The genome-wide association studies are prepared in order to identify genomic regions that affect the variation of phenotypic traits, with the ultimate goal of detecting the causal mutations (Miller et al., 2011). Moreover, the estimation of LD can be applied to the detection of chromosomal regions influenced by selection, and managing genetic resources and diversity. The patterns of LD, reflecting the non-random association of alleles at two or more loci, can provide the insight into the

evolutionary history of a population and can be used to estimate several population parameters including the ancestral effective population size (García-Gómez et al., 2012). Generally, several measures (D' and r^2) have been utilized for determination of LD between marker pairs of loci.

The effective population size (N_e) reflects the loss of genetic variations, accumulation of mutations, effectiveness of selection, and the rate of inbreeding. If the pedigree data are not available, the LD analysis can provide an alternative view to the estimation of N_e . Similarly to most of the other methods of effective population size estimation, the approach based on LD assumes a closed population and randomly recombining loci (Flury et al., 2010). The estimation of N_e can help in predicting genetic variation loss and rate of increase in inbreeding and also provides useful information about evolutionary history of populations. The trend of N_e in livestock populations may give

* Supported by the Slovak Research and Development Agency, contract No. APVV-0636-11 and No. APVV-14-0054.

relevant information for genetic diversity monitoring and helps explain the observed extent of genetic variation in a population from a retrospective point of view (Wang, 2005).

In this study genomic data was used in order to describe the extent of LD within three genetically distant cattle breeds and estimate the effective population size over autosomal genome as one of the most significant indicators of genetic diversity loss as well as the rate of inbreeding increase and genetic drift. Alongside pedigree data, the effective population size estimation based on whole genome scan can provide a more precise view to the state of genetic diversity in populations, mainly to the management of selection scheme and breeding program.

MATERIAL AND METHODS

For assessing the level of LD and subsequently effective population size, three different cattle populations were used. The sources of genomic DNA (semen samples) were obtained from in total 19 Pinzgau, 30 Brown Swiss, and 30 Nelore bulls. All analyzed animals were genotyped in commercial lab using bovine genotyping arrays that included ~50 K SNPs (Illumina, Inc., San Diego, USA). The total number of SNPs on the Illumina BovineSNP50 BeadChip was different within each of analyzed groups. The SNP arrays consisted of 54 609 (Pinzgau), 54 001 (Brown Swiss), and 48 734 SNPs (Nelore). The quality control of genotyping data was performed according to Purcell et al. (2007) for each dataset of analyzed groups separately. In the first step all SNPs with unknown genomic position and SNPs localized on sex chromosomes were excluded. Secondly, based on quality control criteria for remaining SNPs in datasets,

any SNPs with call rate under 0.95, with more than 10% missing genotypes, monomorphic SNPs, and SNPs with minor allele frequency (MAF) under 0.05, and deviation from Hardy-Weinberg equilibrium with limit of 0.001 were filtered out.

In this study, only r^2 that was considered as better interpretable measure for LD was calculated (Qanbari et al., 2010). Only adjacent syntenic loci marker pairs were used to assess the extent of LD. Pairwise r^2 values of SNPs that passed quality control criteria were estimated in the analyzed group of animals independently of each other using SNP & Variation Suite (Version 7.6.8 Win 64; Golden Helix, Bozeman, USA – www.goldenhelix.com). The physical position of SNPs localized on used bovine genotyping arrays can be found in the official Illumina web site (www.illumina.com). SNP pairs were evaluated for distance bins by 50 kb intervals up to 2000 kb and mean r^2 was computed for each interval to determine and illustrate the LD decay in relation to physical distance between markers (Fig. 1). The relationship between LD (r^2) and effective population size (N_e) was estimated according to the approach described by Sved (1971) under the assumption of mutations absence:

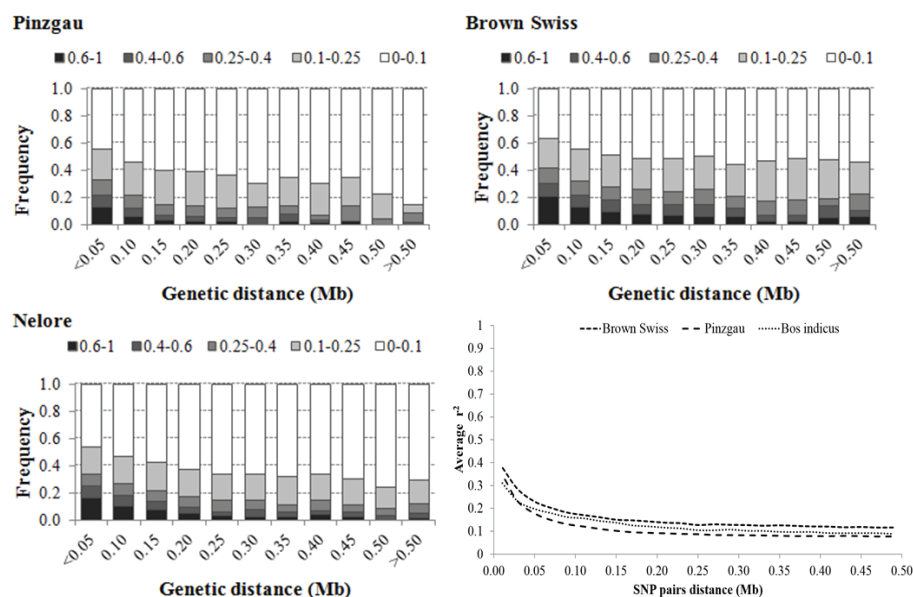
$$E(r^2) = \frac{1}{1 + 4N_e c}$$

where:

c = genetic distance between two loci (in Morgans)

The genetic distance was derived from the inter-marker distance between two considered SNPs (1 Mb ~ 1 cM). According to Umari, Tapio (2011) only adjacent syntenic SNPs with values of $0.01 < r^2 < 0.99$ were used in the calculation of N_e . Finally, historical effective population size $N_e t$ at generation t ($t = \frac{1}{2} c$) was plotted.

Fig. 1. Comparison of the marker pair proportions within different r^2 levels among the analyzed groups within 50 kb distance bins and past effective population size over time based on linkage disequilibrium data



RESULTS

The total number of SNPs that passed the above filtering criteria and has been included into the final LD analysis depends on the used SNP genotyping array and group of animals. After quality control processes of remaining SNPs in datasets for the following analyses of LD 76.08%, 65.13%, and 41.97% SNPs were selected in Pinzgau, Brown Swiss, and Nelore datasets, respectively. The observed SNP distances between usable marker pairs varied from 60.19 (Pinzgau) to 123.74 kb (Nelore).

The MAF across groups was in the range from 0.23 (Nelore) to 0.27 (Pinzgau). The MAFs showed almost the same distribution within different frequency classes in Pinzgau and Brown Swiss bulls (Fig. 2), which probably reflected the different SNPs

composition of used Bovine50K BeadChips and also its application on Nelore cattle.

The extent of LD was significantly different among cattle breeds and varied also within each autosome. All possible adjacent SNP pairs produced 41 512 (Pinzgau), 35 134 (Brown Swiss), and 20 414 (Nelore) pair-wise LD values across all autosomes. The average r^2 values between adjacent syntenic markers were 0.20 for Pinzgau (0.17 BTA29 – 0.23 BTA6), 0.27 for Brown Swiss (0.20 BTA23 – 0.34 BTA6), and 0.19 for Nelore cattle (0.15 BTA27 – 0.27 BTA14). The higher proportion of loci in complete LD was found in Pinzgau cattle (0.26%) compared to the Brown Swiss (0.15%) or Nelore (0.17%). Within each evaluated group the highest proportion of SNPs pairs with $r^2 < 0.3$ was found. The comparable percentages of marker pairs with $r^2 \geq 0.3$ were observed in Pinzgau

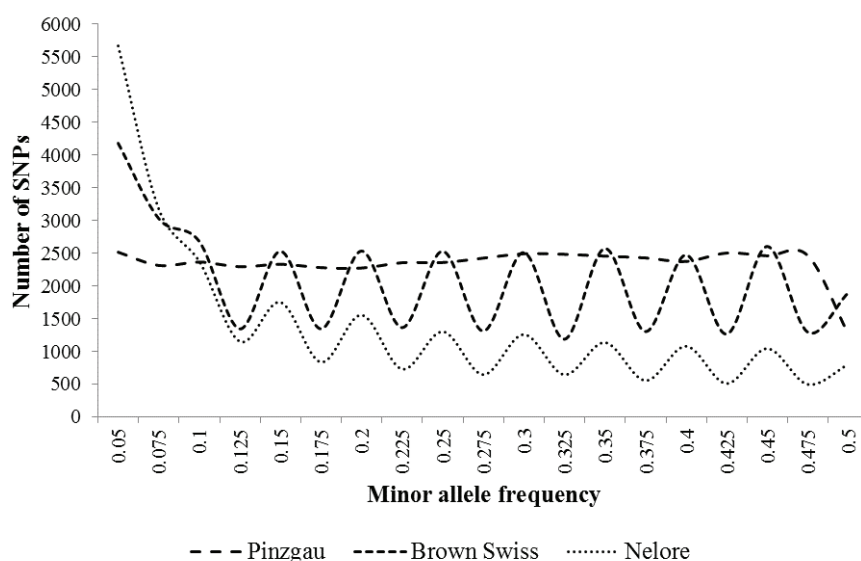
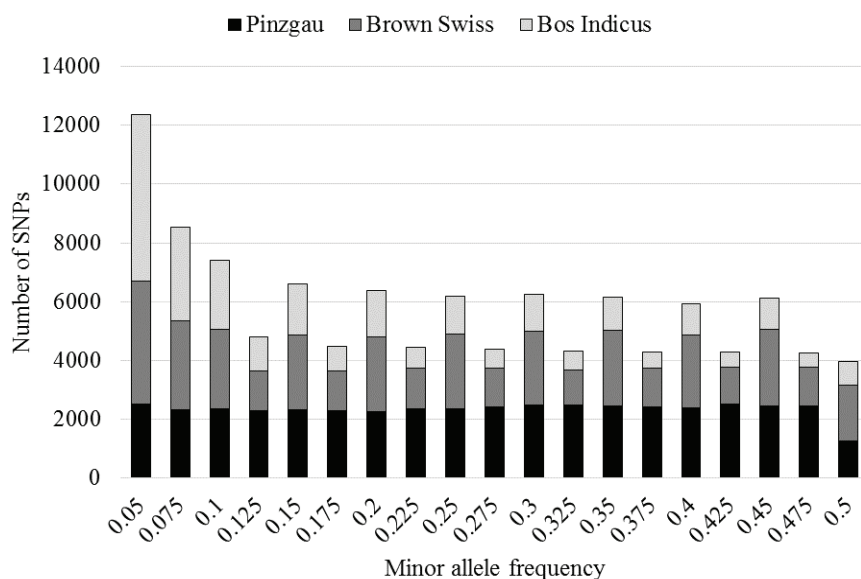


Fig. 2. Minor allele frequency distribution for total SNPs



(22.45%) and Nelore cattle (21.51%), while in the Brown Swiss bulls up to 31.49% of the SNP pairs showed the r^2 higher than 0.3. Fig. 1 illustrates the average r^2 values for SNP pairs separated by 50 kb intervals and by the level of observed r^2 (0–0.1, 0.1–0.25, 0.25–0.4, 0.4–0.6, 0.6–1.0).

Estimation of the effective population sizes using the formula of Sved (1971), assuming no mutations or selection, can lead to difficulties in handling values that are in the limits of the parameter space, because if $r^2 = 0$, the estimate of N_e is infinite, and if $r^2 = 1$, it is also zero (Uimari, Tapio, 2011). For this reason only r^2 values between 0.01 and 0.99 were used in the evaluation of effective population size in our study. The N_e estimation of analyzed cattle breeds during the past is illustrated in Fig. 3. In all populations, in relation to the increase of genetic distances, the decline of r^2 values and also of N_e was found. The average r^2 values in distance bins (50 kb) up to 2000 kb decreased from 0.22 to 0.01 in Pinzgau, from 0.27 to 0.05 in Brown Swiss, and from 0.24 to 0.03 in Nelore cattle. Based on the genomic data, the recent effective population size of Pinzgau bulls was around 172 animals, of Brown Swiss approximately 232 animals, and of Nelore around 240 animals. In the evaluation of past effective population size (500 generations ago) a lower value was found for Brown Swiss (1290) compared to the Pinzgau cattle (1992) and Nelore (1539).

DISCUSSION

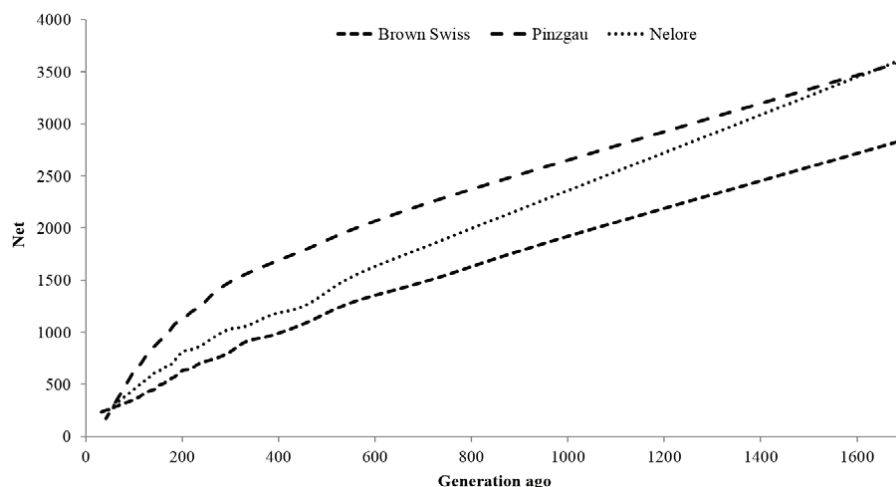
The observed SNP distances between usable marker pairs for LD analysis were slightly higher in comparison with other studies estimating the LD level in different cattle breeds (Flury et al., 2010; Zhu et al., 2013). The usable part of SNPs within each group was not equally distributed on autosomes. The higher proportion of polymorphic SNPs across all autosomes was found for taurine cattle breeds, which was due especially

to the application of genotyping array produced for taurine bovids on evolutionary closer indicine cattle breed (Nelore).

In our study only SNPs with MAF > 0.05 were included. For the estimation of LD the high-frequency SNPs are preferable, because the LD between markers with low MAF is biased upwards (Qanbari et al., 2010). Moreover, the Illumina bovine genotyping array was optimized with respect to uniform SNP spacing and distribution of minor allele frequency primarily to taurine cattle breeds. One possibility how to eliminate or reduce the impact of ascertainment bias is the use of HD genotyping array. O'Brien et al. (2014) observed only negligible impact of MAF changes on the LD estimation in taurine and indicine cattle.

The highest r^2 values were found for BTA6 (taurine cattle) and BTA14 (indicine cattle). Arias et al. (2009) showed the decrease of recombination distance across bovine genome with the length of chromosome which means that the rate of recombination increases with chromosome length. This simply indicated that LD will extend for shorter distances on longer chromosomes and therefore that longer chromosomes will have lower LD than shorter chromosomes (Lee et al., 2011). The observed r^2 levels across autosomes for the analyzed groups were comparable with results published for other dairy cattle breeds (Qanbari et al., 2010; O'Brien et al., 2014). The decay of LD in relation to the increasing of genetic distance was observed within all r^2 levels and evaluated cattle groups similarly as in the recently published studies (Flury et al., 2010; Qanbari et al., 2010; Zhu et al., 2013). The most rapid decrease was found mainly in the first five distance bins up to 0.25 Mb. Generally, the LD across the long distances in the genome is considered as a signal of the recent population history and the LD observed for the short distances reflect the historical effective population size (Hayes et al., 2003). Therefore the strength of LD level at different genetic distances between loci may be used to assess the current and past effective population size (Hill,

Fig. 3. Estimated historical effective population size over time from linkage disequilibrium data



1981). Generally, the observed N_e values indicated that the level of usable animals in populations can be considered as sufficient and the populations as endangered. One of the reasons for the observed decrease of effective population size in recent generations across the analyzed populations can be the application of intensive selection in each cattle breeding program.

CONCLUSION

The results of this study clearly showed that the evaluation of LD extent through genotyping data can be beneficial not only for the detection of QTL in relation to the analysis of phenotypic variation across individuals, but also may be a very effective tool for estimation of N_e . Moreover, several studies assessing the effective population size based on LD evaluation suggested that the genomic information can provide a more precise analysis of population genetic parameters. The estimation of N_e can present relevant data that may help explain the genetic variability in populations also from a retrospective point of view and can enable us to predict the genetic variability loss mainly in the managing of breeding programs in small livestock populations.

REFERENCES

- Arias JA, Keehan M, Fisher P, Coppieters W, Spelman R (2009): A high density linkage map of the bovine genome. *BMC Genetics*, 10, 1471–2156. doi: 10.1186/1471-2156-10-18.
- Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Simianer H, Hanotte O, Rieder O (2010): Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics*, 127, 339–347. doi: 10.1111/j.1439-0388.2010.00862.x.
- García-Gámez E, Sahana G, Gutiérrez-Gil B, Arranz JJ (2012): Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics*, 13, 43. doi: 10.1186/1471-2156-13-43.
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003): Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13, 635–643. doi: 10.1101/gr.387103.
- Hill WG (1981): Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, 38, 209–216. doi: 10.1017/S0016672300020553.
- Lee SH, Cho YM, Lim D, Kim HC, Choi BH, Park HS, Kim OH, Kim S, Kim TH, Yoon D, Hong SK (2011): Linkage disequilibrium and effective population size in Hanwoo Korean cattle. *Asian-Australasian Journal of Animal Sciences*, 24, 1660–1665. doi: 10.5713/ajas.2011.11165.
- Miller J, Poissant J, Kijas J, Coltman D (2011): A genome-wide set of SNPs detects population substructure and long range disequilibrium in wild sheep. *Molecular Ecology Resources*, 11, 314–322. doi: 10.1111/j.1755-0998.2010.02918.x.
- O'Brien AMP, Mészáros G, Utsunomiya YT, Sonstegard TS, Garcia JF, Van Tassell CP, Carvalheiro R, da Silva MVB, Sölkner 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 MA, Bender D, Maller J, Skalp P, De Bakker PL, Daly MJ, Sham PC (2007): PLINK: a tool set for whole genome association and population-based linkage analysis. *The American Journal of Human Genetics*, 81, 559–575. doi: 10.1086/519795.
- Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H (2010): The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics*, 41, 346–356. doi: 10.1111/j.1365-2052.2009.02011.x.
- Shin DH, Cho KH, Park KD, Lee HJ, Kim H (2013): Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. *Asian-Australasian Journal of Animal Sciences*, 26, 1672–1679. doi: 10.5713/ajas.2013.13320.
- Sved JV (1971): Linkage disequilibrium and homozygosity of chromosome segments in finite population. *Theoretical Population Biology*, 2, 125–141. doi: 10.1016/0040-5809(71)90011-6.
- 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: 10.2527/jas.2010-3249.
- Wang J (2005): Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1395–1409. doi: 10.1098/rstb.2005.1682.
- Zhu M, Zhu B, Wang YH, Wu Y, Xu L, Guo LP, Yuan ZR, Zhang LP, Gao X, Gao HJ, Xu SZ, Li Y (2013): Linkage disequilibrium estimation of Chinese beef Simmental cattle using high-density SNP panels. *Asian-Australasian Journal of Animal Sciences*, 26, 772–779. doi: 10.5713/ajas.2012.12721.

Corresponding Author:

Doc. Ing. Radovan Kasarda, Ph.D., Slovak University of Agriculture, Faculty of Agrobiological and Food Resources, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic, phone: +421 376 414 292, e-mail: radovan.kasarda@uniag.sk
