Effective Population Size and Genomic Inbreeding in Slovak Pinzgau Cattle

Nina MORAVČIKOVÀ (✉)
Radovan KASARDA
Veronika KUKUČKOVÀ
Ondrej KADLEČIK

Summary

The aim of this study was to evaluate the level of genomic inbreeding and effective population size in the Slovak Pinzgau population based on molecular – genetic analysis. The genotyping data was obtained from in total 152 animals (37 sires, and 115 cows) representing the active Pinzgau population in Slovakia. All of animals have been genotyped using Illumina BovineSNP50 BeadChip V2 and after quality control the final dataset was composed of 41,738 autosomal loci. The inbreeding coefficient (F_{ROH}) was expressed as the length of the genome covered by runs of homozygosity (ROH) divided by length of the autosomal genome covered by all SNPs (2.5 Gb). Across both groups of sires and cows ROH segments greater than 4 Mb ($F_{ROH > 4\ Mb}$) cover in average 2.22 % of the genome, whereas inbreeding estimates > 16 Mb ($F_{ROH > 16\ Mb}$) achieved 0.81 % that signalized recent inbreeding in analysed population. The historical and recent effective population sizes were estimated based on the relationship between the extent of linkage disequilibrium and effective population size. The estimates of historical effective population size showed linear decrease within each of analysed group. A decrease of 7.81 individuals per generation has been observed. The predicted current N_e across all of animals (30.29) clearly confirmed the endangered status of Slovak Pinzgau population and indicated the need for constant monitoring to increase population size without reduction of genetic diversity due to inbreeding.

Key words

cattle, genome-wide SNP data, effective population size, inbreeding, runs of homozygosity

Department of Animal Genetics and Breeding Biology, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia
✉ e-mail: nina.moravcikova@uniag.sk

Received: May 15, 2017 | Accepted: July 31, 2017

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

Introduction

Effective population size (N_e) and inbreeding coefficient (F) are among the most important parameters in conservation genetics. The effective population size is defined as the number of reproducing individuals, bred in an idealized population in which all individuals are of the same sex and selfing is permitted, and that leads to the same decrease of genetic diversity than the population being studied (Falconer, 1996). The estimation of N_e can help in prediction of loss of genetic variation and rate of increase in inbreeding and also provides useful information about evolutionary history of populations (Sargolzaei et al., 2008; Leroy et al., 2013; Curik et al., 2014). In addition, the knowledge about trend of N_e development in livestock populations provide relevant information for the monitoring of genetic diversity and helps to explain the observed extent of genetic variation in population from a retrospective point of view (Flury et al., 2010). The pattern of historical N_e in livestock populations can increase our understanding of the impact of selective breeding strategies on the genetic variation within the framework of population genetics (Shin et al., 2013).

Estimating N_e has been subject to much research over the last 80 years. Traditionally, effective population size and inbreeding coefficients have been estimated from pedigree records. Classical inbreeding and N_e estimates rarely work well in real populations as they are mostly based on inaccurate pedigree records or, in case of N_e estimation, on robust demographic parameters that do not completely recognise the history of the population (bottlenecks, preferential mating or population subdivision) (Curik et al., 2014; Barbato et al., 2015).

The dense marker chips that have been developed for livestock with the main purpose of increasing selection responses, can be also used for obtaining genome-wide estimates of effective population size and inbreeding. Genomic estimates are expected to be more accurate than pedigree-based estimates because they reflect the actual percentage of the genome that is homozygous (inbreeding) or the actual percentage of the genome shared by two individuals (coancestry), whereas pedigree-based estimates are only expectations of such percentages. Moreover, genomic estimates are able to capture relationships due to very distant common ancestors that pedigree-based estimates ignore (Keller et al., 2011). On the basis of the above the aim of this study was to estimate the effective population size and level of genomic inbreeding based on high-trough SNP data in Slovak Pinzgau cattle.

Material and Methods

Genotyping and quality control of data

The sampled population of Slovak Pinzgau cattle covered living animals: 37 sires representing active breeding bulls (19 animals) and AI doses deposited in reproduction centres (18 animals), 35 dams (nucleus cows) and 80 cows. All of animals have been genotyped in commercial lab for 54906 SNPs using Illumina BovineSNP50 BeadChip V2. The quality control of genotyping data was performed by PLINK (Purcell et al., 2007). Markers assigned to unmapped regions or with unknown chromosomal position according to the latest bovine genome assembly (UMD 3.1) and SNPs positioned to sex chromosomes were removed. The subsequent quality control of genotyping data has been conducted to exclude any SNPs with call rate lower than 90%, minor allele frequency lower than 0.05 and HWE limit of $1x10^{-5}$. The final dataset was composed of 41,738 autosomal loci covering overall length 2.5 Gb of the genome.

Genomic inbreeding

The inbreeding coefficient (F_{ROH}) was characterized as the length of the genome present in runs of homozygosity (ROH) divided by specified length of the autosomal genome covered by SNPs (2.5 Gb) according to McQuillan et al. (2008) using PLINK software (Purcell et al., 2007). The level of F_{ROH} in three different length categories (> 4Mb, > 8Mb, and > 16 Mb) was determined to differentiate between ancient and recent inbreeding. According to Ferenčaković et al. (2013b) the following criteria have been used to characterize the ROH segments: (i) the minimum number of SNPs included in the ROH segments was fixed to 15; (ii) the minimum length of ROH was set to 1 Mb; (iii) minimum density of one SNPs on every 100 kb; (iv) maximum gap between consecutive SNPs of 1 Mb; (v) one heterozygous call was allowed for length >16 Mb. In addition, one missing call was allowed for length >4 Mb, 2 for >8 Mb and 4 for >16 Mb. For each group and ROH length category the total number of ROH detected, the average length of ROH (in Mb) and the sum of all ROH segments by animals have been calculated.

Effective population size

In this study, the relationship between variance in linkage disequilibrium (LD) and N_e was used to infer ancestral and recent effective population sizes. The N_e was estimated using SNeP software (Barbato et al., 2015) that allows the estimation of N_e trends across generation using SNP data that corrects for sample size, phasing and recombination rate based on the formula (Corbin et al., 2012):

$$N_{T(t)} = (4f(c_t))^{-1} \left(E\left[r^2_{adj} \mid c_t\right]^{-1} - \alpha\right)$$

where N_t is the effective population size t generations ago calculated as $t = (2f(c_t))^{-1}$ (Hayes et al., 2003), c_t is the recombination rate (here approximated by the physical distance 1 cM~1 Mb), r^2_{adj} is the LD value adjusted for sample size and α is a correction for the occurrence of mutations (α=2.2 as recommended by Corbin et al. 2012). Under the assumption of the constant linear growth of N_e with the time expressed in past generations (Hayes et al., 2003), the historical effective population size was expressed as a function of time and physical genetic distance between two loci. The current N_e was predicted based on the linear regression performed on estimates obtained for the past generations (N_{eLD10} to N_{eLD60}) as described Kukučková et al. (2017).

Results and discussion

The summary statistics of F_{ROH} within each of analysed group is listed in table 1. For each individual three inbreeding coefficients were calculated based on the ROH of different length ($F_{ROH > 4 Mb}$, $F_{ROH > 8 Mb}$, and $F_{ROH > 16 Mb}$). Different ROH inbreeding coefficients are expected to have differently remote common ancestors (Curik et al., 2014). Ferenčaković et al. (2013a) showed that the ROH segments of 2 – 4 Mb long, representing the 25 – 12.5 generations from common ancestor, correspond most to IBD segments from the past usually not able to be captured with the available pedigree information, although they might also contain some ROHs that were identical by state

Table 1. Summary statistics of the genomic inbreeding coefficients calculated from ROH with different size (> 4 Mb, > 8 Mb, and > 16 Mb) within each of analysed group

Inbreeding coefficient	Mean	Standard deviation	Lower 95% CI	Upper 95% CI	Range
Cows					
$F_{ROH > 4 Mb}$	0.024	0.022	0.019	0.029	0.000 – 0.133
$F_{ROH > 8 Mb}$	0.016	0.020	0.012	0.020	0.000 – 0.131
$F_{ROH > 16 Mb}$	0.008	0.016	0.005	0.012	0.000 – 0.105
Dams					
$F_{ROH > 4 Mb}$	0.017	0.016	0.012	0.023	0.000 – 0.072
$F_{ROH > 8 Mb}$	0.011	0.013	0.006	0.015	0.000 – 0.062
$F_{ROH > 16 Mb}$	0.007	0.012	0.003	0.012	0.000 – 0.048
Sires					
$F_{ROH > 4 Mb}$	0.023	0.019	0.017	0.030	0.000 – 0.072
$F_{ROH > 8 Mb}$	0.016	0.015	0.011	0.021	0.000 – 0.048
$F_{ROH > 16 Mb}$	0.009	0.011	0.005	0.012	0.000 – 0.033

without being IBD. In contrast, very long runs represent recent inbreeding (16 Mb segments are expected mean after ≈ 3 generations), so part of autozygosity that is due to more distant common ancestors is not covered with them. Across all of groups ROH segments greater than 4 Mb ($F_{ROH > 4 Mb}$) cover in average 2.22 % of the genome, whereas inbreeding estimates greater than 16 Mb achieved 0.81 % that signalized recent inbreeding in analysed population. Generally, the lowest number of ROH, length of ROH segments and resulting inbreeding level were found for group of dams. Our study showed lower level of genomic inbreeding within Slovak Pinzgau population compared to the Austrian Pinzgau cattle. Ferenčaković et al. (2013b) and Kukučková et al. (2017) reported overall higher level of F_{ROH} for Austrian sires ($F_{ROH>4Mb}$ = 0.037). In comparison to the pedigree data Kadlečík et al. (2008) reported for population of Slovak Pinzgau cattle the average inbreeding at level 3.08 %. Higher number of inbreed animals was observed in the evaluated population compared to their ancestors, but the mean value of F_x decreased continuously from value of 13.28 % in the 3rd generation of ancestors to 2.48 % in base population. Similarly, Pavlík et al. (2013) found an increasing trend of inbreeding coefficient with increasing number of traced generations taken into account. Moreover, they reported the higher inbreeding level in dairy compared to the beef Slovak Pinzgau population. The average increase in inbreeding was 0.21 % in dairy and 0.04 % in beef population.

The current N_e within each of analysed animal group were predicted based on the linear regression performed on estimates obtained for the past generations (from 10 to 60 generation ago) to summarize the extent of linkage disequilibrium. Under the assumption of the constant linear growth of N_e with the time expressed in past generations, the historical effective population size has been expressed as a function of time and physical genetic distance between loci. Generally the larger distances refer to N_e estimates which are closer to the current generation (Hayes et al. 2003). The estimates of historical effective population size (N_{eT}) showed linear decrease within each of group (Figure 1). The predicted current N_e across all of analysed animals was 30.29 with a 95% confidence interval ranging from 28.95 to 33.46. The obtained linear regression function $N_{eT}=30.285+7.81*t$ with coefficient of determination $R^2 = 0.996$ was comparable with previously published results by Kukučková et al. (2017) which similarly confirmed the endangered status of Slovak Pinzgau population at the moment. Pavlík et al. (2014) reported based on the pedigree data for dairy and beef Slovak Pinzgau populations the value of N_e on the level 122.45 and 809.40, respectively. But they also stated that the effective population size is presumably overestimated in Slovakia, especially in the beef population due to lower pedigree depth.

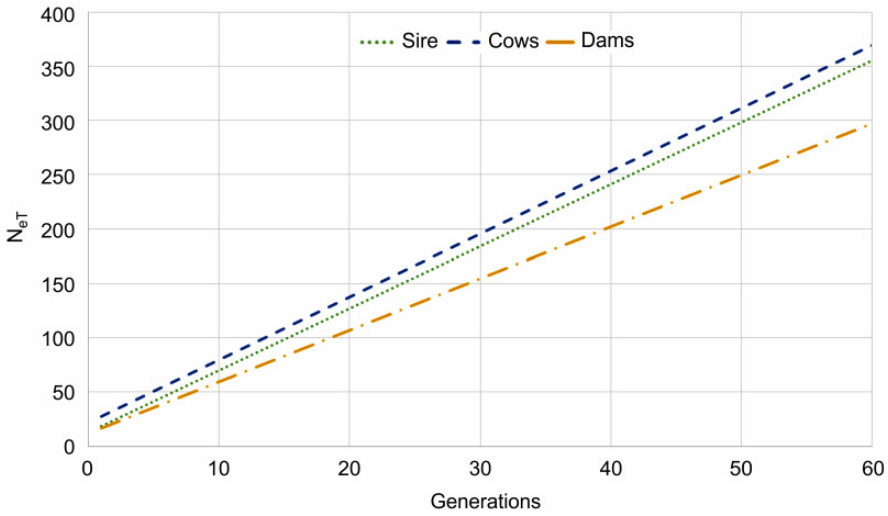

Figure 1. The estimation of N_e trends across generation within each of analysed group based on the linear regression

Conclusion

The analysis of molecular inbreeding coefficient based on the ROH segments covering genome indicated the presence of recent inbreeding in analysed population. Across all of groups ROH segments greater than 4 Mb ($F_{ROH > 4 Mb}$) cover in average 2.22 % of the genome, whereas inbreeding estimates > 16 Mb ($F_{ROH > 16 Mb}$) achieved 0.81 %. The group of dams showed lower number of ROH, length of ROH segments as well as inbreeding level compared to the sires and cows. The estimates of historical effective population size indicated the linear decrease within each of analysed group. The predicted current N_e across all of animals (30.29) clearly demonstrated the endangered status of Slovak Pinzgau population that was previously described based on both pedigree and genomic information. The results of this study reflect the need for constant monitoring to increase population size without reduction of genetic diversity due to inbreeding. In addition, this study will contribute to the conservation management strategy of Pinzgau cattle in Slovakia.

References

Boichard D., Maignel L., Verriér E. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. Gen Sel Evol 29: 5-23

Barbato M., Orozco-terWengel P., Tapio M., Bruford M.W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet 6:109

Corbin L.J., Liu A.Y.H., Bishop S.C., Wooliams J.A. (2012). Estimation of historical effective population size using linkage disequilibrium with marker data. J Anim Breed Genet 129(4): 257-270

Curik I., Ferenčaković M., Sölkner J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest Sci 166: 26-34

Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I., Sölkner J. (2013a). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J Anim Breed Genet 130: 286-293

Ferenčaković M., Solkner J., Curik I. (2013b). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet Sel Evol 45(1): 42

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. J Anim Breed Genet 127: 339-347

Falconer D.S., Mackay T.F.C. (1996). Introduction to Quantitative Genetics. Longman Group Ltd., Harlow

Hayes B.J., Visscher P.M., McPartlan H.C., Goddard M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res 13: 635-643

Kadlečík O., Kasarda R., Mészáros G. (2008). Inbreeding in pure-bred Slovak Pinzgau dual-purpose cattle population. Archiva Zootechnica 11(2): 21-28

Keller M.C., Visscher P.M., Goddard M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 89: 237-249

Kukučková V., Moravčíková N., Ferenčaković M., Simčič M., Mészáros G., Sölkner J., Trakovická A., Kadlečík O., Curik I., Kasarda R. (2017). Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives. Conserv Genet 10.1007/s10592-017-0935-9

Leroy G., Mary-Huard T., Verrier E., Danvy S., Charvolin E., Danchin-Burge C. (2013). Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. Gen Sel Evol 45: 1

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A., Farrington S., Rudan P., Hayward C., Vitart V., Rudan I., Wild S., Dunlop M., Wright A., Campbell H., Wilson J. (2008). Runs of homozygosity in European populations. Am J Hum Genet 83: 359-372

Pavlík I., Kadlečík O., Kasarda R., Šidlová V., Žitný J. (2013). Comparison of genetic diversity in dual-purpose and beef Pinzgau population. Acta fytotechn zootechn 16(3): 69-73

Pavlík I., Sölkner J., Kadlečík O., Kasarda R., Mészáros G., Fuerst Ch., Fuerst-Waltl B. (2014). Joint genealogical analysis as a tool for diversity evaluation in Pinzgau cattle populations. Archiv Tierzucht 57(14): 1-12

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

Sargolzaei M., Schenkel F.S., Jansen G.B., Schaeffer L.R. (2008). Extent of linkage disequilibrium in Holstein cattle in North America. J Dairy Sci 91: 2106-2117

Shin D.H., Cho K.H., Park K.D., Lee H.J., Kim H. (2013). Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. Asian-Australas J Anim Sci 26: 1672-1679

acs82_17