Assessment of genetic diversity and analysis of pedigree completeness in three horse breeds in Slovakia

E. Hazuchová[1], R. Kasarda[1†], O. Kadlečík[1], I. Pavlík[1], J. Žitný[1], N. Moravčíková[1], M. Horný[2]

[1]*Slovak University of Agriculture, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic;*
[2]*National Stud Topoľčianky, Parková 13, 951 93 Topoľčianky, Slovak Republic*

SUMMARY

The aim of the study was to assess the diversity based on pedigree information in three horse populations. The pedigree information was available from The National Stud Topoľčianky, Slovakia. The whole analysed populations consisted of (reference populations in brackets) 537 (58) Lipizzan horses, 1984 (45) Shagya Arabian horses and 1105 (31) Hucul horses. The equivalent complete generations ranged from 5.48 for the Lipizzan, to 7.35 for the Shagya Arabian horses. The average of inbreeding coefficient ranged from 1.16 for the Lipizzan to 5.11 for the Hucul horses. The highest average individual increase of inbreeding ΔF_i = 0.82 % and the average individual relatedness coefficient AR = 7.82 % was found for the Hucul horses. Therefore the lowest value of effective population size N_e = 58.73 was found for Hucul horses. The assessment populations were derived from 90 to 255 founders. The effective number of founders ranged from 23 to 83, the effective number of ancestors from 16 to 23, while founder's genome equivalent ranged from 5.45 to 11.23. Only a small number of ancestors were needed to explain half of the genetic diversity in the studies populations (from 6 for the Hucul and Shagya Arabian to 9 for the Lipizzan). Results will be used in genetic management of breeding work in Slovakian horse breeds and monitoring of parameters characterizing genetic diversity and their development, as well.

Keywords: average relatedness, diversity, effective number of founders, founders, inbreeding, pedigree completeness

INTRODUCTION

Genetic diversity has been defined as the variety of alleles and genotypes present in a population and this is reflected in morphological, physiological and behavioural differences between individuals and populations (Frankham et al.

[†] Contact person e-mail: Radovan.Kasarda@uniag.sk

2002). One way to describe genetic variability and its evolution across generations is through the analysis of pedigree information. Genealogical information would yield extensive parameters to assess the actual levels of diversity, and therefore we preferred to assess the state of variation. An important step in the analysis of the population structure is a calculation of the effective population size, which quantifies genetic drift and the rate of inbreeding in a population (Teegen et al. 2009). The rate of inbreeding is a common criterion to monitor changes of genetic diversity. However, this method is very sensitive to incomplete pedigree information. Another complementary approach is to analyze the probabilities of gene origin as numbers of founders, effective number of founders, and ancestors. Mating of relatives was a routine method in the improvement of animals and their productive qualities. The increased homozygoteness, which was the main genetic effect from inbreeding, determined the appearance of recessive genes effect, the creation of a similar genotypic and phenotypic related groups and a differentiation of the based population in to different lines and families (Falconer and Mackay, 1989). Evaluation of genetic diversity using pedigree information in horse populations provided Cervantes et al. (2008), Siderits et al (2013), Hamann and Distl (2008). The aim of this study was to assess the diversity based on pedigree information in three horse populations in Slovakia using parameters on probability of identity by descent and gene origin.

MATERIAL AND METHODS

Assessment of the genetic diversity of the Lipizzan, Shagya Arabian and Hucul populations was made on the basis of pedigree information from the database The National Stud Topoľčianky. We assess the genetic diversity of individuals within stud. Basic characteristic of the assessed files are given in Table 1. For data processing software SAS V9.2 was used. For monitoring of genetic variability in populations was used ENDOG v4.8 software (Gutiérrez and Goyache, 2005). The genealogical information was completed to maximize number of ancestral generations used in the analysis.

The pedigree completeness level of the population was assessed by using 3 parameters:

1. Maximum number of generation traced (GenMax) is numbers of generations' separating the individual from its furthest ancestor,

2. Number of fully traced generation (GenCom) is defined as those separating the offspring of the furthest generation where the ancestors of the individual are known,

3. Equivalent complete generations (GenEqu) is computed as the sum over all known ancestors of the terms computed as the sum of $(1/2)^n$ where n is the

number of generations separating the individual to each known ancestor (Gutiérrez et al., 2005).

Table 1 Basic characteristics of the assessment files

		Lipizzan		Shagya Arabian		Hucul	
	Birth years	1922 – 2007		1979 – 2007		1988 - 2008	
	n	58		45		31	
RP	Sex	M	F	M	F	M	F
	n	9	49	8	37	9	22
	Birth years	1991 - 2007		1904 - 2007		1903 – 2008	
	n	537		1984		1105	
WP	Sex	M	F	M	F	M	F
	n	210	327	610	1374	283	822

RP – reference population, WP – whole population, M – male, F – female

Generation interval

The generation interval id defined as the average age of parents at the birth of their progeny kept for reproduction (James, 1977). This parameter was computed for the 4 pathways (father – son, father – daughter, mother – son and mother – daughter) and average values of generation interval.

Measures based on the probability of identity by descent

The *inbreeding coefficient (F_i)* - is defined as the probability that an individual has two identical alleles by descent (Gutiérrez, et al., 2005).

The *individual increases in inbreeding* (ΔF_i) are computed as: $\Delta F_i = 1 - {}^{t-1}\sqrt{1 - F_i}$, where F_i is the individual coefficient of inbreeding and *t* is the equivalent complete generations (Gutiérrez and Goyache, 2005).

The effective population size N_e

The effective population size is defined as the number of individuals in an ideal population that would give rise to the same variance of gene frequencies or the same rate of inbreeding as observed in the actual breeding population (Falconer and Mackay, 1996). Realized of effective population size $\overline{Ne}$ (Cervantes et al., 2008) can be computed from ΔF that can be computed by averaging ΔF_i of the *n* individuals included in a given reference subpopulation as: $\overline{Ne} = \dfrac{1}{2\overline{\Delta F}}$

Note that this way of computing effective population size is not dependent on the whole reference population mating policy but on the mating carried out throughout the pedigree of each individual (Gutiérrez et al., 2008).

The *average relatedness coefficient (AR)* of each individual is defined as the probability that an allele randomly chosen from the whole population in

the pedigree belongs to a given animal. AR can then be interpreted as the representation of the animal in the whole pedigree regardless of the knowledge of its own pedigree (Gutiérrez and Goyache, 2005).

Measures based on probability of gene origin

All animals with both parents unknown are regarded as founders in this study. *Effective number of founders* (f_e), is defined as the number of equally contributing founders that would be expected to produce the same genetic diversity. This is computed as; $f_e = \dfrac{1}{\displaystyle\sum_{k=1}^{f} q_k^2}$, where q_k is the probability of gene origin of the *k* ancestors (Gutiérrez and Goyache, 2005). When founders contribute unequally, the effective number of founders is smaller than the actual number (Sørensen et al., 2005). *Effective number of ancestors* (f_a) is minimum number of ancestors, explaining the complete genetic diversity of a population. This is computed as $f_a = \dfrac{1}{\displaystyle\sum_{j=1}^{a} q_j^k}$, where q_j is marginal contribution of an ancestor *j*, which is the genetic contribution made by an ancestor that is not explained by other ancestors chosen before (Gutiérrez and Goyache, 2005). The effective number of ancestors (f_a) is less than the f_e and the comparison of both numbers can be used to find the bottlenecks that have occurred from the founders to the present population (Boichard et al., 1997).
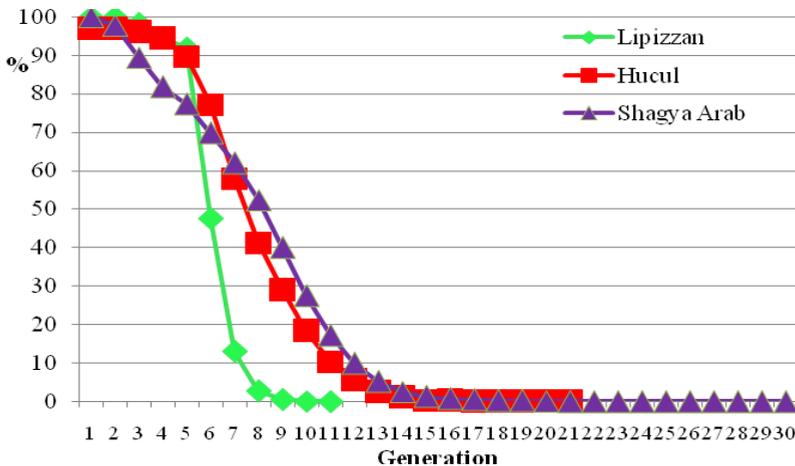

Figure 1 Ratio of known ancestors (%) per parental generation

RESULTS AND DISCUSSION

Values of pedigree completeness are given in Figure 1. Pedigrees the first 5 generations are more or less complete for all assessment of breeds. Proportion of known ancestors decreased fewer than 50 %, after 6 generations in Lipizzan, after 8 generations in Hucul, and after 9 generations in and Shagya Arabian. In the study of Zechner et al. (2002) was found in the 10 generation about 90 % of ancestors known for Lipizzan breed. The pedigree data were used to calculate other pedigree completeness parameters as well. The average values of the maximum number of generations traced and equivalent complete generations of ancestors for the Lipizzan, Hucul and Shagya Arabian Gen Max 8.15±1.28; 17.70±3.71; 26.93±3.71 and GenEqu 5.48±0.41; 7.16±1.59; 7.35±1.14, respectively. The results for Hucul horses were similar as results presented by Pjontek et al. (2012) for the Hucul, where they found GenMax = 17.54; GenCom = 4.29; GenEqu = 7.10, but lower for Shagya Arabian and Lipizzan, where they found for the Shagya Arabian and Lipizzan GenMax 34.82 and 24.56, GenCom for the 5.58 and 5.90, GenEqu 9.56 and 10.25, respectively. The greatest of values for Shagya Arabian and Lipizzan in the study Pjontek et al. (2012) due to fact they assessed whole population in Slovakia including individuals from private studs.

Table 2. Average generation intervals in references populations

Breed		Lipizzan	Shagya Arabian	Hucul
Generation interval	$\bar{x}$	12.23	12.81	12.17
	s	4.65	5.65	4.87

For monitoring of genetic diversity in the Lipizzan horses were used only 5 generation of ancestors. In this a reason we found the lowest values of pedigree completeness for this breed. The average values of generation interval for three horse breeds in Slovakia are very similar and they are shows in the Table 2. Our results were slightly longer than as results presented by Pjontek et al. (2012), where they found average generation interval in range 11.14 for Hucul to 12.27 for Shagya Arabian horses and results presented by Valera et al. (2005) for Andalusian horse population. The long generation intervals usually reported in the horse are basically dependent on its use which is not compatible with pregnancy and breeding life (Valera et al. 2005).

Table 3, figure 2 and figure 3 show the results of assessment based on the probability of identity by descent of assessed populations. The highest average value of inbreeding and increase in inbreeding was found for the Hucul breed, having values F_i= 5.11 % and ΔF_i= 0.82 %. Our result were lower as in the study Pjontek et al. (2012), which could be due to a lower level of pedigree completeness. These parameters are very sensitive to incomplete pedigree

information. In the case of pedigree information are incomplete will be results of inbreeding and increase in inbreeding coefficients underestimated. Therefore, the results of characteristics based on the probability of identity by descent were lowest for Lipizzan horses. Using five generation pedigrees for calculating inbreeding coefficient leads to values around 2 % (Zechner et al., 2002). The inbreeding coefficients based on five generations pedigrees are in the study Moureaux et al. (1996).

Table 3 Characteristics based on the probability of identity by descent in reference populations

		Breed								
		Lipizzan			Shagya Arabian			Hucul		
		M	F	RP	M	F	RP	M	F	RP
F_i	$\bar{x}$	1.93	1.02	1.16	4.25	3.81	3.89	5.29	5.03	5.11
	s	1.42	0.99	1.1	3.28	2.58	2.68	1.56	2.83	2.50
	x_{min}	0.39	0	0	0.99	0	0	3.21	0	0
	x_{max}	4.88	3.90	4.88	9.26	9.41	9.41	8.01	12.75	12.75
ΔF_i	$\bar{x}$	0.46	0.22	0.26	0.67	0.58	0.60	0.84	0.81	0.82
	s	0.34	0.22	0.26	0.50	0.37	0.39	0.22	0.45	0.39
	x_{min}	0.10	0	0	0.19	0	0	0.54	0	0
	x_{max}	1.18	0.94	1.18	1.50	1.60	1.60	1.19	1.88	1.88
AR	$\bar{x}$	3.06	2.99	3.0	3.98	4.20	4.16	7.48	7.97	7.82
	s	0.60	0.42	0.45	1.49	0.71	0.88	1.16	2.20	2.03
	x_{min}	1.86	1.95	1.86	0.11	2.08	1.13	4.01	0.21	0.21
	x_{max}	3.68	3.57	3.68	0.54	5.43	5.49	9.38	10.50	10.50
$\overline{Ne}$		-	-	187.75	-	-	83.18			58.73

F_i- the average individual coefficient of inbreeding; ΔF_i– the average individual increase of inbreeding; AR - the average relatedness coefficient; $\overline{Ne}$ - the effective population size; M – male; F – female

Our results for Lipizzan horses are similar as results of inbreeding coefficients for Trotteur Français (F = 1.86 %) and Anglo – Arab (1.17 %). The effective population size (N_e) has been used as a key parameter in designing strategies for the definition and conservation of endangered animal species (Bijma et al. 2001). We found the N_e ranged from 58.73 for Hucul to 187.75 for Lipizzan horses. Missing parents are supposed to be unrelated to the defined actual breeding population. This can affect the increase in inbreeding and a result, the estimated N_e might be overestimated. The greatest value of N_e for Lipizzan horses due to larger part missing parents and therefore the increase in inbreeding in population was to be underestimated. Required of effective population size for conservation of genetic diversity in the population by Frankham et al. (2002), is 500. In this respect we cannot the assessed of

populations horses in Slovakia described as sustainable. Because a close relationship exists between N_e and all parameters derived from the probabilities of gene origin, conclusions can also be drawn for this for effective number of founder genomes (f_g). The smaller N_e, he grater the decrease in the f_g per unit time (Boichard et al., 1997).
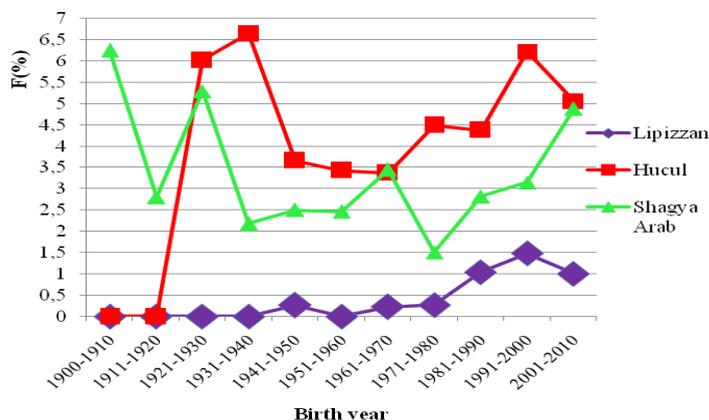

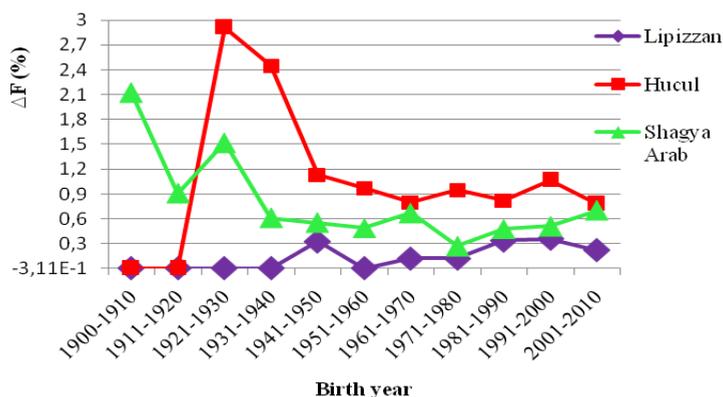Figure 2 Inbreeding trend in three horse breeds in Slovakia

Figure 3 Average increase of inbreeding ΔF in three horse breeds in Slovakia

The parameters characterizing the genetic diversity of the populations in terms of gene origin are given in Table 4. The populations of Shagya Arabian and Lipizzan were derived from 255 and 156 founders, respectively, while populations of Hucul only 90. When founders contribute unequally, the effective number of founders is smaller than the actual number (Sørensen et al. 2005). The decrease of genetic variability assessed through these parameters is reflected in lower values, such as for the effective number of founders (ranged from 28 for the Hucul to 83 for Lipizzan) and effective number of ancestors (ranged from 16 for Hucul and Shagya Arabian to 23 for

Lipizzan). The founder's genome equivalent accounts for the loss of genetic diversity due to unequal founders' contribution and also random genetic drift. It is a more accurate description of variation. All determinations of genetic loss are relative to the genetic diversity in the founder population. The analysis of allelic loss is therefore a determination of the probability that alleles present in the founder population still reside within the descendant population (Lacy, 1989).

Table 4 Characteristics based on the probability of gene origin

Parameters	Breed		
	Lipizzan	Shagya Arabian	Hucul
Number of animals (*n*)	58	45	31
Number of founders (*f*)	156	255	90
Effective number of founders (*f$_e$*)	83	49	28
Effective number of ancestors (*f$_a$*)	23	16	16
Founder's genome equivalent(*f$_g$*)	11.23	6.33	5.45
Number of ancestors explaining 50% genetic variability	9	6	6

Our results for Shagya Arabian were slightly greater as results by presented Cervantes et al. (2008), where they found for Spanish Arab horse f_e = 39.5 and f_a = 13. We found the f_g ranged from 5.45 for the Hucul to 11.23 for Lipizzan. The results of f_e, f_a, f_g for Lipizzan horses in Topolčianky were greater as results presented by Zechner et al. (2002). In comparison to the study Pjontek et al. (2012), which found greater results of f, f_e, f_a, f_g for all three assessed populations, while number of ancestors explaining 50 % of genetic diversity was similar with results found in presented study. Higher f_e than f_a and higher f_g than f_a indicates the occurrence of bottleneck and genetic drift in all assessed populations. Only a small number of ancestors were needed to explain half of the genetic diversity in the studies subpopulations. Thus it is likely that these groups will produce half sibs, which will mate and subsequent generations of descendant will increase the average value of inbreeding (Pjontek et al., 2012).

CONCLUSIONS

Throughout the present analysis we have analyzed the pedigree information of three Slovakian horses' breeds. Higher values of relatedness coefficient in Hucul will reflect improvement of inbreeding coefficient in the next generation. The comparison between the number of founder and effective numbers of founders advert to a decrease of genetic diversity as consequence of unequal contributions of founders. This could happen due to

excessive use of some animals as parents of subsequent generations. The knowledge of population structure gives an overview of the population and is the first step for developing optimal mating and selection strategies for the future. In order to prevent loss of diversity within breed it is needed to monitor and use suitable mating strategy. Using only restricted number of generations (five) leads to uncertainty in estimation of parameters based on pedigree information. Full pedigree information has to be used due to historically based relatedness.

REFERENCES

Bijma, P., Van Arendonk, J.A.M., Wooliams, J.A., 2001. Predicting rate of inbreeding for livestock improvement schemes. J. Anim. Sci. 79, 840-853.

Boichard, D., Maignel, L., Verrier, É., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29, 5-23.

Cervantes, I., Molina, A., Goyarche, F., Gutiérrez, J.P., Valera, M., 2008. Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. Livestock Sience. 113, 24-33.

Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J. P., 2008. Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. J. Anim. Breed. Genet. 125, 301-310.

Falconer, D.S. and Mackay, T.F.C., 1989. Introduction to quantitative genetics. Longman, London and NY, 340 p.

Falconer D.S. and Mackay T.F.C., 1996. Introduction to quantitative genetics 4th edition. Longman Group Ltd, Essex, London.

Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to conversation genetics. Cambrige University Press, 617 p.

Gutiérrez, J. P., Cervantes, I., Molina, A., Valera, M., Goyache, F., 2008. Individual increase in inbreeding allows estimating realized effective sizes from pedigrees. Genet. Sel. Evol. 40, 359-378.

Gutiérrez, J. P., Goyarch, F., 2005. A note on ENDOG: a computer program for Analysis pedigree information. J. Anim. Breed. Genet. 122, 172-176.

Hamann, H. and Dist,l O., 2008. Genetic variability in Hanoverian warmblood horses using pedigree analysis. J. Anim. Sci. 86, 1503-1513.

James, J. W., 1977. A note on selection differentials and generation length when generations overlap. Anim. Prod. 24, 109-112.

Moureaux, S., Verrier, E., Ricard, A., Meriaux, J.C., 1996. Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. Genet. Sel.Evol. 28,83-102.

Pjontek, J., Kadlečík, O., Kasarda, R., Horný, M., 2012. Pedigree analysis in four Slovak endangered horse breeds. Czech J. Anim. Sci. 57, 2, 54-64.

Siderits, M., Baumung, R., Fuerst-Waltl, B., 2013. Pedigree analysis in the German Paint Horse: Genetic variability and the influence of pedigree quality. Livestock Science. 151, 152-157.

Sørensen, A. C., Sørensen, M. K., Berg, P., 2005. Inbreeding in Danish Dairy Cattle Breeds. J. Dairy Sci. 88, 1865-1872.

Teegen, R., Edel, C., Thaller, G., 2009. Population structure of the Trakehner Horse breed. Animal, 3, 1, 6-15.

Valera, M., Molina, A., Gutiérraz, J.P., Gómez, J., Goyache, F., 2005. Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. Livestock Production Science. 95, 57-66.

Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., Brem, G., 2002. Analysis of diversity and population structure in the Lipizzan horse breed based on the pedigree information. Livestock Production Science. 77, 137-146.