

## The genetic structure of indigenous Romanian Hucul horse breed inferred from microsatellite data

S. E. GEORGESCU, MARIA ADINA MANEA, MARIETA COSTACHE

University of Bucharest, Molecular Biology Center, 91-95 Splaiul Independentei, 5 Bucharest, Romania, E

### Abstract

*The existence of the Hucul horse on the Romanian territory has been documented from the very distant past and some of the theories state that this horse originates from a wild mountain horse, similar to the Tarpan. Today, the Hucul is an independent and unique breed and belongs to the protected gene fund of original and primitive animal breeds of FAO. The main objective of this study was the analysis of genetic diversity based on microsatellites in this old horse breed from Eastern Europe. This breed has been comparatively studied to other Romanian breeds: Romanian Sport Horse, Thoroughbred and Arabian. The observed and expected heterozygosity per breed ranged from 0.662 and 0.676 (Hucul) to 0.759 (Thoroughbred) and 0.741 (Romanian Sport Horse), respectively. Hardy-Weinberg equilibrium was tested for all breed-combinations and the exact P-value over all loci and breeds was not significant. Only 12% of the total genetic variability could be attributed to differences among breeds ( $F_{ST}=0.123$ ,  $p<0.01$ ). The global population differentiation tests showed highly significant ( $p<0.01$ ) results for all 12 loci. The Hucul clearly demarcated from the other horse population in the UPGMA tree.*

Keywords: Hucul horse, microsatellite, genetic diversity, clustering.

### Introduction

Over the past few years, the issue of maintaining biodiversity as a major element of environment preservation has been discussed globally. To this end, preserving biodiversity of indigenous species, especially of those of economic interest must represent a relevant aspect in the scientific research activity. For four decades now, FAO (Food and Agriculture Organization) has included in its agenda the problem of preserving, evaluating and using animal genetic resources.

The existence of the Hucul horse on the Romanian territory has been documented from the very distant past and some of the theories state that this horse originates from a wild mountain horse, similar to the Tarpan. It is thought that the area that most shaped present-day Hucul is the North of Bucovina. The Hucul horse is characterized by its perfect adjustment to mountain areas, high resistance to effort, perfect sense of direction and it is sure-footed for mountain walks on narrow paths. Originally bred in the modest and harsh conditions of the Carpathians, Hucul has all the features of a wild horse as it is adapted to low temperature areas with high rainfall in mountain regions (1400 metres altitude). This breed has been isolated and strengthened as a race and it has evolved almost entirely to its present day form by natural selection.

Today, the Hucul is an independent and unique breed. Since 1979, Hucul belongs to the protected gene fund of original and primitive animal breeds of FAO.

In phylogenetic relationships studies among different breed, the molecular techniques are widely used. Microsatellites loci comprise an attractive potential source of information

about population histories and evolutionary processes [11]. Due to their accuracy of typing, the highly polymorphic level and the spread distribution in genome, these markers have been successfully used in parentage and relatedness tests in horses [6, 17] and they are very useful in estimating the genetic distance among different populations [23]. In the last decade microsatellites were used for the analyses of different horse breeds around the world, also including primitive horse breeds: Przewalski horse [7], Spanish Celtic breeds [8], Norwegian breeds [3, 4] and various European and Asian breeds [24, 25].

In this study we investigate the genetic divergence between four horse populations: a Romanian indigenous one (Hucul), a Romanian horse breed still unformed (Romanian Sport Horse) and two breeds which were isolated for a long time (Thoroughbred and Arabian). The Romanian Sport Horse is a horse breed whose formation is in progress and it started in 1963. The Thoroughbred horse was chosen because it is a breed which was isolated for a long time and the Arabian horse was selected due to its influences on most of the European riding horses.

This analysis included the distribution within and between breeds of the observed genetic variation, phylogenetic analysis of breed assignment from microsatellite allele frequencies.

## Materials and Methods

### Sampling and DNA extraction

Fresh blood of 240 individuals from the following horse populations was collected: Thoroughbred (TB), Arabian (AR), Romanian Sport Horse (RSH) and Hucul (HC) (Table 1). The individuals were chosen at random and we avoided closely related animals. The isolation of genomic DNA from fresh blood was performed with Wizard Genomic DNA Extraction Kit (Promega).

**Table 1.** Population size sampled and the total population size, in Romania (2006).

Breed	Number of animals sampled			Population size
	F	M	Total	
Thoroughbred (TB)	35	25	60	219
Arabian (AR)	40	20	60	312
Romanian Sport Horse (RSH)	37	23	60	384
Hucul (HC)	38	22	60	350

### Microsatellite analyses

Amplification of the microsatellite loci was realized by multiplex PCR using StockMarks® for Horses Equine Genotyping Kit (AppliedBiosystems) and it was carried out using a GeneAmp 9700 System (AppliedBiosystems). In our study we included 12 microsatellite markers because only these ones were successfully amplified for all animals tested (Table 2). PCR products were detected by capillary electrophoresis using an ABI Prism 310 DNA Genetic Analyzer (AppliedBiosystems). The size of alleles was determined by using GeneScan-500 LIZ Size Standard and the results were processed with the GeneScan® 3.1.2 and Genotyper® 2.5.2 Softwares (AppliedBiosystems). As a reference we used the DNA control from the kit.

**Table 2.** The panel of microsatellites used in the analysis [22].

Microsatellite loci	Chromosome	Size obtained (bp)			
		HC	RSH	TB	AR
VHL20	30	86-106	86-106	86-100	86-106
HTG4	9	128-138	126-140	124-138	126-130
AHT4	24q14	144-164	144-158	144-160	144-162
HMS7	1	171-181	171-181	171-181	171-181
HTG6	15q26-q27	80-102	80-102	78-102	78-96
HMS6	4	158-168	158-168	154-168	154-168
HTG7	4	118-126	118-126	118-130	118-130
HMS3	9	148-168	148-168	148-172	148-172
AHT5	8	130-144	130-140	126-140	128-140
ASB2	15q21-q23	234-250	238-252	236-264	234-254
HTG10	21	82-106	86-104	86-110	82-110
HMS2	10	216-232	216-236	218-228	216-238

### Statistical analysis

The allelic frequencies, observed and expected heterozygosities ( $H_o$  and  $H_e$ ) were estimated using the CERVUS 2.0 program [18] and the Hardy-Weinberg equilibrium test (HWE) was performed with the GENEPOP 4.0 software [20]. The exact P-values were assessed using the complete enumeration method [15] for loci with fewer than five alleles or using the Markov-Chain algorithm [13] if otherwise (with 1000 dememorization steps for 100 batches and 5000 iteration per batch). The hypothesis that all four horse breeds are significantly distinguishable on the basis of genic and genotypic differentiation was then tested. Afterwards differentiation tests were performed for each locus for each breed to evaluate the significance of genic and genotypic differentiations for all pairs of populations.

For the estimation of difference between populations, we used  $F_{ST}$  [27] indices because the genetic drift is assumed to be the main factor in genetic differentiation among closely related populations or for short-time evolution [19, 26, 23]. Using the FSTAT program [12] we also estimated the  $F_{IS}$  and  $F_{IT}$  indices, for each population.

We used Reynolds' genetic distance ( $D_R$ ) under the assumption that all gene frequencies are changed through genetic drift alone [19].

PHYLIP 3.5 statistical package [10] was used to calculate genetic distances, and to obtain bootstrap procedures and trees. Bootstraps' values were computed over 1000 replicates, and a consensus tree was drawn.

## Results and Discussions

### Level of variation and HWE

A total of 119 different alleles were detected for all the 12 analyzed microsatellites. The entire batch of 12 microsatellites were polymorphic, out of which the most polymorphic is HTG10 with a total of 12 alleles, meanwhile HMS7, with 6 alleles in total, is the least polymorphic (Table 3). The observed and expected heterozygosities and the mean number of alleles (MNA) together with their standard deviations are showed in Table 4. Observed and expected heterozygosities per breed ranged from 0.662 and 0.676 (HC) to 0.759 (TB) and 0.741 (RSH), respectively. HWE was tested for all breed-combinations. The exact P-values over all loci and breeds were not significant. When results were pooled across breeds, three

microsatellites offered a significant deviation: VHL20 ( $p < 0.05$ ) in RSH; AHT5 ( $p < 0.01$ ) in TB and RSH and HTG10 ( $p < 0.01$ ) in TB and AR.

**Table 3.** The number of alleles per locus in each population.

Locus	TB	AR	RSH	HC	Total
VHL20	6	8	8	10	10
HTG4	7	3	4	5	9
AHT4	6	8	7	7	11
HMS7	5	5	6	6	6
HTG6	9	6	6	7	11
AHT5	8	7	5	7	9
HMS6	6	7	5	5	7
ASB2	11	8	8	7	12
HTG10	12	11	7	9	15
HTG7	7	7	3	5	7
HMS3	10	8	7	7	11
HMS2	5	9	8	5	11

#### F-statistics

The  $F_{ST}$  values for each locus are represented in Table 5. The multilocus  $F_{ST}$  values indicate that around 12.3% of the total genetic variation could be due to the breeds' differences and the remaining 87.7% may correspond to differences among individuals. The  $F_{ST}$  values for single loci varies from 0.052 (ASB2) to 0.273 (HTG7). On average, each of the four breeds had a 7% deficit of heterozygotes, whereas the entirety of individuals had an 18.5% deficit of heterozygotes. The global population differentiation tests (genic and genotypic) showed highly significant ( $p < 0.01$ ) result for all 12 loci.

**Table 4.** Observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosities and the mean number of alleles (MNA) over 12 microsatellites in 4 horse populations.

Population	$H_O$	$H_E$	MNA
Thoroughbred (TB)	0.759±0.09	0.720±0.108	8.08±2.71
Arabian (AR)	0.691±0.12	0.738±0.101	7.25±2.005
Romanian Sport Horse (RSH)	0.709±0.064	0.741±0.07	6.16±1.46
Hucul (HC)	0.662±0.135	0.676±0.122	6.58±1.62

Highly significant ( $p < 0.01$ ) genetic differences for all breeds combinations were shown through pairwise tests. The  $F_{ST}$  values ranged from 9.7% for the RSH-HC pair to 1.8% for the AR-HC pair (Table 6).

**Table 5.** F statistical estimates and their significances by locus.

Locus	$F_{ST}$	$F_{IS}$	$F_{IT}$
VHL20	0.058	0.028	0.084
HTG4	0.137	0.048	0.179
AHT4	0.146	-0.018	0.131
HMS7	0.07	-0.038	0.035
HTG6	0.186	0.164	0.32
AHT5	0.085	0.163	0.234
HMS6	0.095	0.027	0.121
ASB2	0.052	0.099	0.146
HTG10	0.118	0.083	0.192

HTG7	0.273	0.139	0.374
HMS3	0.093	0.108	0.191
HMS2	0.16	0.038	0.192
Mean	<b>0.123</b>	<b>0.07</b>	<b>0.185</b>

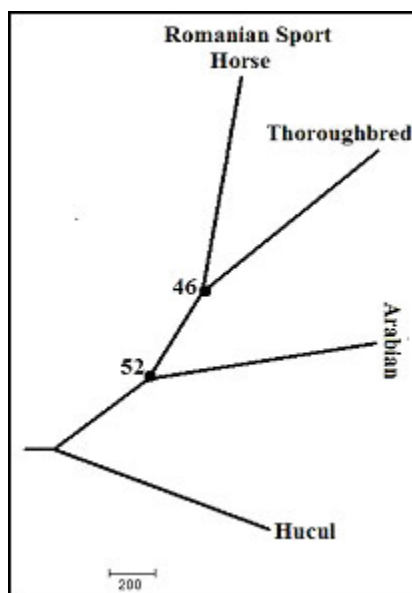
**Table 6.** Fst estimates compared in pairs (above diagonal) and Reynolds's genetic distance (below diagonal).

	TB	AR	RSH	HC
TB		0.0997	0.0878	0.1501
AR	0.108410		0.1116	<b>0.1877</b>
RSH	0.096567	0.119528		0.0979
HC	0.157910	<b>0.194924</b>	0.105775	

### Breed relationships

The  $D_R$  distance ranged from 0.096 to 0.194. The two smallest distances for the pair RSH-TB and the largest distance between HC and AR were calculated (Table 6).

A UPGMA tree based on Reynolds's genetic distance relating the four horse populations is presented in Figure 1. The numbers of the nodes are bootstrapping values for 1000 replicates of the 12 loci genotyped.



**Figure 1.** Phylogenetic tree constructed from Reynolds' genetic distance by UPGMA method. The numbers at the nodes are values for 1000 bootstrap resamplings of the 12 loci genotyped.

The 0.123 values (Table 5) suggest an overall differentiation of 12.3% between breeds. This is comparable with the differentiation observed between other horse breeds: Norwegian horse breed, 12% [3]; German draught horse breeds 11.6% [1], but is greater than that observed between Spanish Celtic horse breeds (8%) [8] and that found among Spanish donkey breeds (3.6%) [2]. This value is higher than that observed in European cattle breeds (10%) [16] and in dogs (9.9%) [14] but is lower compared to the 17% observed in goats [21] and 10-20% observed in humans [9].

The results obtained by the Hardy-Weinberg test demonstrate that all the four horse populations are in equilibrium, without any digressions from it.

The phylogenetic tree obtained using the UPGMA method based on Reynolds' genetic distances show an early and clear divergence of the Hucul breed in comparison to the others

three breeds analyzed. These data confirm that the Hucul breed is an ancient one which has undergone a divergence from the common branch before the Arabian and Thoroughbred. On the other hand, the Romanian Sport Horse, a new breed whose formation is still in progress, is the latest to evolve from the common branch.

As it is observed in Figure 1, the clear-cut divergence of the Hucul from the other three breeds is in agreement with the existing data in the literature regarding its ancient origins. The first written references to the Hucul horses of Bucovina appear in 1603. This horse is considered the most valuable equine material in the area of the Carpathian Mountains and it is the only autochthonous breed surviving over the centuries, while the Dobrudja horse and the well-known Moldavian horse have become extinct in time. The origins of this mountain horse are lost in the passage of the centuries and are quite controversial. Based on the morphologic parameters analysis and on its exterior appearance, the Hucul resembles both the reconstitution of the Tarpan horse [5], as well as the wild Mongol horse, *Equus przewalski*. In this respect, the Hucul horse is attributed a certain phylogenetic closeness to the Oriental horse, but there are sources which consider that the Hucul horse and the Polish Konik horse are related and descend directly from the Tarpan horse.

## Conclusions

In order to accurately determine the controversial origin of the Hucul horse, we need to extend our analyses to include other primitive breeds (Icelandic, Exmoor Horse) or on the Przewalski wild horse. Unfortunately, a comparison of the Hucul with the wild Tarpan horse is impossible to achieve as the last wild Tarpan horse became extinct in 1879 in Ukraine.

Our preliminary results are going to be completed through the mtDNA analysis of the Hucul and through construction of phylogenetic trees based on mtDNA sequences. Comparative analyses, on the level of microsatellites and mitochondrial DNA, between the Hucul horses and other primitive breeds (Polish Konik, Przewalski Horse, Icelandic or Exmoor Pony) could help elucidate the origin and phylogenetic relations of this ancient horse. However, this is the first microsatellite-based genetic diversity study performed on the Hucul population in Romania. These data could be considered a first step forward in confirming the primitive origin of this breed.

## References

1. ABERLE K. S., HAMANN H., DRÖGEMÜLLER C., DISTL O. (2004) Genetic Diversity in German Draught Horse Breeds Compared with a Group of Primitive, Riding and Wild Horses by Means of Microsatellite DNA Markers. *Anim. Genet.* **35**, 270-277.
2. ARANGUREN-MENDEZ J., JORDANA J., GOMEZ M. (2001) Genetic Diversity in Spanish Donkey Breeds Using Microsatellite DNA Markers. *Genet. Sel. Evol.*, **33**, 433-442.
3. BJØRNSTAD G., GUNBY. E., RØED K. H. (2000) Genetic Structure of Norwegian Horse Breeds. *J. Anim. Breed. Genet.*, **117**, 307-317.
4. BJØRNSTAD G., RØED K. H. (2001) Breed Demarcation and Potential for Breed Allocation of Horses Assessed by Microsatellite Markers. *Anim. Genet.*, **32**, 59-65.
5. BOWLING A.T., RUVINSKY A. (2000) *The Genetics of the Horse*. CABI Publishing, New York, USA.
6. BOWLING A. T., EGGLESTON-STOTT M. L., BYRNS G., CLARK R. S., DILEANIS S., WICTUM E., (1997) Validation of Microsatellite Markers for Routine Horse Parentage Testing. *Anim. Genet.*, **28**, 247-252.
7. BREEN M., LINDGREN G., BINNS M. M., NORMAN J., IRVIN Z., BELL K., SANDBERG K., ELLEGREN H. (1997) Genetical and Physical Assignments of Equine Microsatellites-First Integration of Anchored Markers in Horse Genome Mapping. *Mamm. Genome*, **8**, 267-273.

8. CANÓN J., CHECA M. L., CARLEOS C., VEGA-PLÁ J. L., VALLEJO M., DUNNER S. (2000) The Genetic Structure Of Spanish Celtic Horse Breeds Inferred From Microsatellite Data., *Anim. Genet.*, **31**, 39-48.
9. CAVALLI-SFORZA L. L., MENOZZI P., PIAZZA A. (1994) The History And Geography Of Human Genes. Princeton University Press, Princeton, NJ. USA.
10. FELSENSTEIN, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164-166.
11. GOLDSTEIN D.B., SCHLÖTTERER C. (1999) Microsatellites: Evolution and Applications. Oxford University Press, New York, USA.
12. GOUDET, J. (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *J. Hered.*, **86**, 485-486.
13. GUO, S. W. and THOMPSON E. A. (1992) Performing the Exact Test Of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, **48**, 361-372.
14. JORDANA J., PIEDRAFITA J., SANCHEZ A., PUIG P. (1992) Comparative F Statistics Analysis of the Genetic Structure of Ten Spanish Dog Breeds. *J. Hered.*, **83**, 367-374.
15. LOUIS E. J., DEMPSTER E. R. (1987) An Exact Test for Hardy-Weinberg and Multiple Alleles. *Biometrics*, **43**, 805-811.
16. MacHUGH D. E., LOFTUS R. T., CUNNINGHAM P., BRADLEY D. G. (1998) Genetic Structure Of Seven European Cattle Breeds Assessed Using 20 Microsatellite Markers. *Anim. Genet.*, **29**, 333-340.
17. MARKLUND S., ELLEGREN H., ERIKSSON S., SANDBERG K., ANDERSSON L., (1994) Parentage Testing And Linkage Analysis In The Horse Using A Set Of Highly Polymorphic Microsatellites. *Anim. Genet.*, **25**, 19-23.
18. MARSHALL T. C., SLATE KRUUK L., PEMBERTON J. M., (1998) Statistical Confidence For Likelihood-Based Paternity Inference In Natural Populations. *Mol. Ecol.*, **7**, 639-655.
19. REYNOLDS J., WEIR B.S., COCKERHAM C. (1983) Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics*, **105**, 767-779.
20. ROUSSET F. (2007) GENEPOP007: A Complete Reimplementation of the GENEPOP Software for Windows and Linux. *Mol. Ecol.*, **8**, 103-106.
21. SAITBEKOVA N., GAILLARD C., OBEXER-RUFF G., DOLF G. (1999) Genetic Diversity in Swiss Goat Breeds Based on Microsatellite Analysis. *Anim. Genet.*, **30**, 36-41.
22. SHIUE Y. L., BICKEL L.A., CAETANO A.R., MILLON L.V., CLARK R.S., EGGLESTON M.L., MICHELMORE R., BAILEY E., GUERIN G., GODARD S., MICKELSON J.R., VALBERG S.J., MURRAY J.D., BOWLING A.T. (1999) A Synteny Map of The Horse Genome Comprised Of 240 Microsatellite And RAPD Markers. *Anim. Genet.*, **30**, 1-9.
23. TAKEZAKI N., NEI M. (1996) Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA. *Genetics*, **144**, 389-399.
24. TOZAKI T., TAKEZAKI N., HASEGAWA T., ISHIDA N., KUROSAWA M., TOMITA M., SAITOU N., MUKOYAMA H. (2003) Microsatellite Variation in Japanese and Asian Horses and Their Phylogenetic Relationship Using a European Horse Outgroup. *J. Hered.*, **94**, 374-380.
25. VILÁ C., LEONARD J.A., GÖTHRESTRÖM A., MARKLUND S., SANDBERG K., LIDÉN K., WAYNE R. K., ELLEGREN H. (2001) Widespread Origins of Domestic Horse Lineages. *Science*, **291**, 474-477.
26. WEIR B. S. (1990) Genetic Data Analysis. Sinauer Associates, Sunderland, USA.
27. WEIR B. S., COCKERHAM C. C. (1984) Estimating F-Statistics for the Analysis Of Population Structure. *Evolution*, **38**, 1358-1370.