https://doi.org/10.17221/2/2018-CJAS

Using nuclear microsatellite data to trace the gene flow and population structure in Czech horses

Lenka Putnová[1], Radek Štohl[2], Irena Vrtková[1]*

[1]Laboratory of Agrogenomics, Department of Morphology, Physiology and Animal Genetics, Faculty of Agronomy, Mendel University in Brno, Brno, Czech Republic

[2]Department of Control and Instrumentation, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

**Corresponding author: putnova@email.cz*

Citation: Putnová L., Štohl R., Vrtková I. (2019): Using nuclear microsatellite data to trace the gene flow and population structure in Czech horses. Czech J. Anim. Sci., 64: 67–77.

Abstract: Based on a data set comprising 2879 animals and 17 nuclear microsatellite DNA markers, we propose the most comprehensive in-depth study mapping the genetic structure and specifying the assignment success rates in horse breeds at the Czech population scale. The STRUCTURE program was used to perform systematic Bayesian clustering via the Markov chain Monte Carlo estimation, enabling us to explain the population stratification and to identify genetic structure patterns within breeds worldwide. In total, 182 different alleles were found over all the populations and markers, with the mean number of 10.7 alleles per locus. The expected heterozygosity ranged from 0.459 (Friesian) to 0.775 (Welsh Part Bred), and the average level reached 0.721. The average observed heterozygosity corresponded to 0.709, with the highest value detected in the Czech Sport Pony (0.775). The largest number of private alleles was found in *Equus przewalskii*. The population inbreeding coefficient F_{IS} ranged from –0.08 in the Merens to 0.14 in the Belgian Warmblood. The total within-population inbreeding coefficient was estimated to be moderate. As expected, very large genetic differentiation and small gene flow were established between the Friesian and *Equus przewalskii* (F_{ST} = 0.37, Nm = 0.43). Zero F_{ST} values indicated no differences between the Czech Warmblood–Slovak Warmblood and the Czech Warmblood–Bavarian Warmblood. A high level of breeding and connectivity was revealed between the Slovak Warmblood–Bavarian Warmblood, Dutch Warmblood–Oldenburg Horse, Bavarian Warmblood–Dutch Warmblood, and Bavarian Warmblood–Oldenburg Horse. The breeds' contribution equalled about 6% of the total genetic variability. The overall proportion of individuals correctly assigned to a population corresponded to 82.4%. The posterior Bayesian approach revealed a hierarchical dynamic genetic structure in four clusters (hot-blooded, warm-blooded, cold-blooded, and pony). While most of the populations were genetically distinct from each other and well-arranged with solid breed structures, some of the entire sets showed signs of admixture and/or fragmentation.

Keywords: admixture; breed stratification; gene migration; genetic variation; horse; individual assignment

The genetic structure of a population, defined as the community of individuals sharing a common gene pool, has evolved through the action of past selective forces on the genes controlling variability. Model-based clustering has been developed to detect the underlying population structure in a collection of individuals genotyped with multiple markers. An advantage of the analysis implemented

Supported by the Ministry of Agriculture of the Czech Republic (Project No. QH92277) utilising the institutional development support of Mendel University in Brno, and by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LO1210).

https://doi.org/10.17221/2/2018-CJAS

with the broadly employed STRUCTURE software (Pritchard et al. 2000) consists in its ability to estimate the proportion of the genome of an individual that belongs to each inferred population (admixture). Understanding how genetic variation is distributed within and among populations is important and useful to the development of breeding strategies and conservation programs.

In the Czech Republic, horse breeding embodies a well-established process cultivated for centuries. After 1990, however, the original system within this domain experienced a gradual decline: The free access to external markets, privatisation of breeding services, and influx of foreign breeding companies resulted in an immense expansion of foreign genetics. Since 1997, the number of horses in the Czech Republic has been constantly rising; compared to the statistics of 2003, the sum has doubled in the last 10 years, increasing by nearly 40 000 animals. At present, about 86 000 horses are registered in the Czech Republic, with the total having grown by 6000 over the last four years. The relevant regulation in force, namely, Act No. 154/2000 Coll., on the selection, breeding, and data recording of farm animals (Animal Breeding Act; http://eagri.cz/public/web/mze/), as amended by Act No. 130/2006 Coll., constitutes the legal basis for horse breeding in the discussed region and respects the EU framework in that it recognises breeders associations as the carriers of stud books and entrusts them with the formulation and guaranteeing of breeding programmes. The currently most widespread horse breed in the Czech Republic is the Czech Warmblood, accounting for 28% of horses bred in the country. The second most numerous breed then consists in the Thoroughbred, whose use is not limited to horse racing only. Another subset holding breeds with numerous individuals comprises the Slovak Warmblood, Welsh breeds, and cobs. Popular cold-blooded horses include also the Silesian Noriker and Czech-Moravian Belgian Horse (http://www.aschk.cz). Hypervariable microsatellites, a common universal DNA genetic marker suitable for multiple genetic applications, were extensively employed in horse genotyping to provide an insight into the diversity patterns of horse breeds worldwide (Leroy et al. 2009; Van de Goor et al. 2011; Barcaccia et al. 2013; Berber et al. 2014; Gupta et al. 2014; Putnova et al. 2018).

In this article, we undertook genetic structure analyses to yield information on the Czech equine population dynamics and investigated the individual assignment success. Using a large-scale data set of 2879 animals and utilising nuclear microsatellite DNA polymorphisms at 17 loci, (*i*) STRUCTURE was tested to explain the population stratification within breeds raised in the Czech Republic, and (*ii*) GENECLASS allowed us to estimate the breed identification. Further, the level of the intra- and inter-population genetic diversity across the 43 populations sampled was established.

MATERIAL AND METHODS

Population samples and DNA isolation. This study exploited 2879 samples collected between 2004 and 2014, covering 43 populations. Unrelated individuals from various locations, studs, and horse stables in the Czech Republic can be roughly grouped as including 27 populations of warmblooded horses, 4 of cold-blooded horses, 11 of pony, and 1 of Przewalski's horse (as the outgroup). The investigated populations are summarised in Table 1. Out of the 43 populations, three breeds were represented by individuals from geographically distinct regions, namely, the Camargue, Murgese, and Icelandic horse from France, Italy, and Iceland, respectively. We have performed multiple tests in Czech equine populations since 2001; however, the relevant individuals show different numbers of genetic markers because 17-marker typing was introduced by the authors in a routine analysis of 2004. Thus, the individuals typed before the indicated period exhibited genetic profiles with only 12 loci and were excluded. The total genomic DNA was extracted from the blood, hair roots, or semen straws, using QIAamp® Blood/Tissue kit (Qiagen, USA) according to the manufacturer's protocol.

Genotype determination. The co-amplification of 17 nuclear microsatellite DNA markers (*AHT4*, *AHT5*, *ASB2*, *ASB17*, *ASB23*, *CA425*, *HMS1*, *HMS2*, *HMS3*, *HMS6*, *HMS7*, *HTG4*, *HTG6*, *HTG7*, *HTG10*, *LEX3*, and *VHL20*) was performed in one multiplex PCR reaction (GeneAmp® PCR System 9700; Applied Biosystems, USA), using the commercially available StockMarks® for Horses 17-Plex Genotyping Kit (Applied Biosystems) pursuant to the manufacturers' instructions. Each 1 µl of the PCR product and 0.5 µl of the GeneScan™ 500 LIZ® Size Standard dye (Life Technologies, USA) was loaded in 11.5 µl of Hi-Di™ Formamide (Life Technolo-

https://doi.org/10.17221/2/2018-CJAS

gies). The samples were then denatured at 95°C for 5 min and cooled down for another 5 min. The genotype scoring was performed on an ABI PRISM 310™ Genetic Analyzer (Applied Biosystems). An alphabetical nomenclature was used for the allele size designation in accordance with the ISAG.

Statistical analysis. The allele frequencies, total number of alleles (TNA), number of private alleles (NPA), number of rare alleles (NRA), polymorphic information content (PIC), observed heterozygosity (H_O), and expected heterozygosity (H_E) were calculated across the given loci and populations by means of the Excel Microsatellite Toolkit (Park 2001).

The Hardy–Weinberg equilibrium (HWE) test was performed with the GENEPOP 4.2.1 software (Rousset 2008) via the heterozygote deficit for each locus in each population. The score test (or U test) was employed, and the genotypic linkage disequilibrium (LD) was studied. The exact *P*-values were obtained using a Markov chain Monte Carlo (MCMC) simulation of 10 000 dememorisation steps, 500 batches, and 5000 iterations. A global test across the loci and populations, namely, the multisample score test, was also applied to the hypothesis of the heterozygote deficit. When conducting multiple tests, the levels of significance were adjusted via the sequential Bonferroni technique. The program FSTAT 2.9.3 (Goudet 2001) was exploited to estimate the *F*-statistics, and the statistical significance of the F_{IS} was tested based on randomisation. The amount of long-term gene flow (Nm) between the populations was indirectly estimated based on the F_{ST} value (Slatkin and Barton 1989). The probabilistic assignment of individuals to the predefined populations was tested through the Bayesian statistical-based approach of Rannala and Mountain (1997) implemented in GeneClass 2.0 h (Piry et al. 2004).

Finally, the genetic structure of the populations was assessed using Bayesian clustering methods in STRUCTURE 2.3.4 (Pritchard et al. 2000). An admixture model and a correlated allele frequency model with default parameters were used to analyse the data set. In order to estimate the number of clusters (*K*), ten independent runs were performed for a *K* between 2 and 43 in all the equine populations, with the burn-in period of 50 000 steps; the procedure was followed by 150 000 MCMC iterations. Using STRUCTURE HARVESTER v0.6.94 (Earl and von Holdt 2012), we calculated the mean likelihood, lnP(*K*); the standard deviation for each value of *K*; and Δ*K*, the second order rate of change of the likelihood with respect to *K* (Evanno et al. 2005). The HARVESTER also generated the in-files to be used with the CLUMPP application employed. The CLUMPP version 1.1.2 (Jakobsson and Rosenberg 2007) was utilised to maximise the measure of similarity in the *Q*-matrices of the ten replicates, where the highest average pairwise similarity is defined as *H* (Nordborg et al. 2005) and/or *H'*. The algorithm LargeKGreedy was used for the attempts to find the optimal alignment of the replicates and to test all possible input orders of the runs. The population outputs, or *Q*-matrices, were visualised in DISTRUCT (Rosenberg 2004).

RESULTS AND DISCUSSION

Genetic variations and breed differentiation. In the present study, 182 different alleles were found over all the populations and markers, with the mean number of 10.7 alleles per locus. The microsatellites were all well-amplified and polymorphic. There exist 20 private alleles (i.e. those found only in a single population among the data set of the populations). A large number of private alleles were found in the PRZ (5.97%; *ASB17* – allele *U*, *ASB17* – allele *Y*, *ASB23* – allele *M*, *HMS3* – allele *T*), ICE (2.44%; *ASB17* – allele *D*, *ASB2* – allele *F*, *HMS2* – allele *U*), and HUC (2.26%; *ASB17* – allele *W*, *AHT5* – allele *H*, *CA425* – allele *H*). The mean frequency of the private alleles was 0.083 in the whole data set; however, the value calculated for the PRZ equalled 0.239. Indeed, we observed high frequencies (> 0.05) for 8 out of the 20 private alleles. The most extreme examples consisted in the private alleles *M* (*ASB23*), *A* (*HTG10*), and *F* (*ASB2*), which reached the frequencies of 0.692, 0.191, and 0.127 in the PRZ, FUR, and ICE populations, respectively. Rare alleles (i.e. those found at frequencies < 0.05) were observed across the populations and amounted to 29% of all detected alleles (Table 1). The highest NRAs were observed in the QH and PH. The microsatellites *CA425*, *HMS1*, *HTG4*, *HTG6*, and *HTG7* showed very high allele frequencies (above 0.40) across all the tested equine breeds. The intrapopulation variation parameters at the microsatellite loci are presented in Supplementary Tables S1, S2 and S3 in Supplementary Online Material (SOM). Significant

Czech Journal of Animal Science, 64, 2019 (2): 67–77

https://doi.org/10.17221/2/2018-CJAS

Table 1. Summary statistics and proportion of assignment as determined in GENECLASS and STRUCTURE programs for 43 populations

Population	Code	Sample size	TNA	MNA	NPA	NRA	PIC	F_{IS}	H_O	H_E	Correctly assigned by GENECLASS[1]		Proportion of assignment; STRUCTURE cluster ($K = 4$)[2]			
											n	success rate (%)	1 – Cold-blooded	2 – Pony	3 – Hot-blooded	4 – Warm-blooded
Akhal-Teke	AKT	100	101	5.94	0	20	0.641	0.005	0.681	0.684	98	98.00	0.0449	0.0883	**0.7620**	0.1049
Andalusian	AND	15	91	5.35	0	12	0.656	0.021	0.712	0.727	15	100.00	0.2789	0.1346	0.2050	***0.3815***
Appaloosa	APP	100	132	7.76	0	49	0.711	0.015	0.739	0.750	70	70.00	0.1120	0.1280	0.1886	**0.5714**
Arabian	ARAB	100	113	6.65	0	42	0.646	0.056	0.656	0.695	91	91.00	0.0240	0.0230	**0.8825**	0.0705
Bavarian Warmblood	BAV	24	108	6.35	0	25	0.700	0.031	0.734	0.756	16	66.67	0.0737	0.0514	0.1506	**0.7242**
Belgian Warmblood	BEL	11	89	5.24	0	18	0.671	0.135	0.653	0.750	10	90.91	***0.3504***	0.0912	0.1005	***0.4580***
Camargue	CAM	100	140	8.24	2	46	0.733	0.075	0.709	0.767	67	67.00	0.1974	0.1618	***0.3125***	***0.3283***
Czech-Moravian Belgian Horse	CMB	100	112	6.59	1	28	0.662	0.010	0.694	0.701	96	96.00	**0.7603**	0.0919	0.0502	0.0976
Czech Sport Pony	CSP	74	133	7.82	0	38	0.729	–0.014	0.775	0.765	48	64.86	0.1241	***0.3423***	***0.3614***	0.1722
Czech Warmblood	CZW	100	131	7.71	0	49	0.714	0.014	0.745	0.756	32	32.00	0.0599	0.0659	0.1724	**0.7018**
Dutch Warmblood	KWPN	39	113	6.65	0	31	0.708	0.027	0.736	0.756	25	64.10	0.0802	0.0449	0.1679	**0.7070**
Fjord	FJO	82	105	6.18	0	29	0.646	0.012	0.683	0.691	81	98.78	0.0491	**0.8713**	0.0351	0.0445
Friesian	FRI	100	66	3.88	0	16	0.411	0.003	0.458	0.459	100	100.00	**0.9651**	0.0115	0.0124	0.0110
Furioso	FUR	22	103	6.06	1	22	0.686	0.000	0.745	0.745	18	81.82	0.0540	0.0566	***0.3122***	**0.5772**
Haflinger	HAF	100	109	6.41	0	39	0.639	–0.006	0.689	0.685	98	98.00	**0.8461**	0.0685	0.0427	0.0427
Hannoverian	HAN	56	117	6.88	0	43	0.700	0.019	0.733	0.747	36	64.29	0.0811	0.0885	0.2454	**0.5850**
Holsteiner	HOL	82	116	6.82	0	39	0.687	0.018	0.719	0.733	61	74.39	0.0721	0.0896	0.1163	**0.7220**
Hucul	HUC	100	133	7.82	3	49	0.709	0.000	0.748	0.748	98	98.00	0.1162	**0.5355**	0.2169	0.1314
Icelandic Horse	ICE	80	123	7.24	3	40	0.679	0.034	0.694	0.718	79	98.75	0.0300	**0.9131**	0.0290	0.0279
Irish Cob	IRI	43	112	6.59	0	36	0.693	0.011	0.731	0.739	42	97.67	**0.5088**	0.2427	0.1105	0.1380
Kinsky Horse	KIN	100	116	6.82	0	38	0.688	0.007	0.727	0.732	73	73.00	0.0560	0.0479	0.1502	0.7460
Lipizzan	LIP	10	70	4.12	0	0	0.576	–0.073	0.717	0.671	10	100.00	0.1021	0.1574	**0.5725**	0.1681
Merens	MER	15	83	4.88	0	21	0.582	–0.081	0.708	0.657	14	93.33	**0.5878**	0.0752	0.1209	0.2161
Miniature Horse	MIN	24	114	6.71	0	38	0.679	0.006	0.724	0.728	19	79.17	0.1174	**0.7473**	0.0632	0.0721
Mini Appaloosa	MAPP	15	90	5.29	1	11	0.670	0.000	0.736	0.736	15	100.00	0.0819	**0.8254**	0.0475	0.0452
Moravian Warmblood	MOW	36	102	6.00	0	28	0.678	0.029	0.708	0.729	31	86.11	0.0395	0.0470	0.1727	**0.7408**
Murgese	MUR	51	116	6.82	1	33	0.704	0.024	0.731	0.749	50	98.04	***0.3679***	0.1895	0.2473	0.1953
Noriker	NOR	83	123	7.29	1	44	0.696	–0.016	0.751	0.739	61	73.49	**0.6955**	0.1233	0.0941	0.0870

https://doi.org/10.17221/2/2018-CJAS

Table 1 to be continued.

Population	Code	Sample size	TNA	MNA	NPA	NRA	PIC	F_{IS}	H_O	H_E	Correctly assigned by GENECLASS[1]		Proportion of assignment; STRUCTURE cluster ($K = 4$)[2]			
											n	success rate (%)	1 – Cold-blooded	2 – Pony	3 – Hot-blooded	4 – Warm-blooded
Oldenburg Horse	OLD	21	107	6.29	0	26	0.704	0.038	0.732	0.761	11	52.38	0.0967	0.0702	0.2241	**0.6091**
Old Kladruber Horse	KLA	100	102	6.00	0	22	0.657	0.021	0.686	0.701	97	97.00	0.2845	0.0888	**0.5510**	0.0757
Paint Horse	PH	99	136	8.00	1	51	0.711	0.001	0.750	0.750	67	67.68	0.0693	0.1233	0.2471	**0.5603**
Przewalski's Horse	PRZ	13	67	3.94	4	7	0.546	−0.019	0.645	0.633	13	100.00	0.0535	**0.9037**	0.0240	0.0188
Quarter Horse	QH	100	134	7.88	0	52	0.705	0.001	0.744	0.745	74	74.00	0.0691	0.1092	0.2202	**0.6015**
Selle Français	SF	12	98	5.76	0	16	0.686	0.122	0.668	0.757	10	83.33	0.1487	0.0330	0.0571	**0.7612**
Shagya	SHA	100	106	6.24	0	27	0.649	0.028	0.675	0.695	97	97.00	0.0250	0.0256	**0.9063**	0.0431
Shetland Pony	SHP	95	115	6.76	1	39	0.654	0.035	0.675	0.699	87	91.58	0.0401	**0.8563**	0.0446	0.0590
Silesian Noriker	SNOR	100	116	6.82	0	39	0.665	−0.021	0.727	0.711	85	85.00	**0.7493**	0.0802	0.0944	0.0761
Slovak Warmblood	SLW	100	127	7.47	0	46	0.717	0.039	0.729	0.759	39	39.00	0.0711	0.0645	0.1430	**0.7214**
Standardbred	STA	100	116	6.82	0	37	0.673	0.008	0.709	0.714	89	89.00	0.0477	0.0796	0.1916	**0.6812**
Thoroughbred	THO	100	96	5.65	0	24	0.660	−0.001	0.711	0.711	88	88.00	0.0170	0.0312	0.0613	**0.8904**
Trakehner	TRA	32	107	6.29	0	32	0.681	0.016	0.723	0.734	23	71.88	0.0638	0.0506	0.1568	0.7288
Welsh Part Bred	WPB	87	137	8.06	0	43	0.741	0.029	0.753	0.775	55	63.22	0.1921	0.2088	***0.3940***	0.2050
Welsh Pony and Cob	WPC	58	127	7.47	1	40	0.708	0.038	0.720	0.748	51	87.93	***0.3174***	***0.3554***	0.2033	0.1239
Total		2879	4752		20	1385					2340					
Mean			110.51	6.50	0.47	32.21	0.671	0.016	0.709	0.721		82.38				

TNA = total number of observed alleles, MNA = mean number of alleles/loci, NPA = number of private alleles (i.e. those found in only a single population within the entire data set), NRA = number of rare alleles (i.e. those found at frequencies < 0.05), PIC = polymorphic information content, F_{IS} = population inbreeding coefficient, H_O = observed heterozygosity, H_E = expected heterozygosity

[1]individual assignment as calculated by GENECLASS using the Bayesian method (approach of Rannala and Mountain, 1997)

[2]proportion of individuals assigned to each cluster based on the STRUCTURE analysis at $K = 4$. The largest assignment proportion for each population is shown in bold; the rates of 30–50% are in italics

https://doi.org/10.17221/2/2018-CJAS

($P < 0.05$) deviations of the HWE were observed only in 37 (5.38%) of the 688 autosomal marker-population combinations. Thus, most of the horse populations tested for each locus via the heterozygote deficit showed genotypic frequencies in agreement with the H–W principle, but the HWE was out of imbalanced at more than two loci in the TRA, WPC, SF, and BEL; however, the global test did not reject the HWE, except for the SF (likely due to the small population size). In general, the deviations could be explained by the effects of inbreeding, selective breeding, individual stallions, and random influences. The LD analysis of the overall microsatellite marker combinations in each population revealed that only 1.37% of the total of 5848 combinations exhibited a significant LD, and 0.31% and 0.22% of these were related to the KLA and CAM populations, respectively.

The H_E ranged from 0.459 (FRI) to 0.775 (WPB), and the average level reached 0.721. The average H_O corresponded to 0.709, with the highest value in the CSP (0.775). Regarding the genetic resources, we observed substantial genetic variation, and thus the highest levels of heterozygosity, in the HUC. The H_E and MNA in the Czech HUC (0.75/7.82) correspond to the results published for the Polish HUC (0.73/7.00) by Fornal et al. (2013); in this study, too, no loss of genetic diversity was detected in the two endangered Czech draught horse breeds (the CMB and SNOR), which are now closed, with the use of other stallions not permitted. As regards the ARAB raised in the Czech Republic, genetic diversity has been maintained in spite of the reputed genetic purity and inbreeding practice within the breed. By contrast, the FRI was found to be the least variable and clearly the most inbred population, owing to genetic isolation. Similar results for the FRI and ARAB had been obtained previously (Leroy et al. 2009; Van de Goor et al. 2011).

The population inbreeding coefficient, F_{IS}, ranged from –0.08 (the MER breed) to 0.14 (the BEL breed). The values of the F_{IS} did not differ significantly from zero in any sampled population. The mean total inbreeding coefficient was estimated to be moderate (10.1%). In terms of the *F*-statistics, the overall F_{IS}, F_{IT}, and F_{ST} values equalled, in all the horse populations, 0.016, 0.110, and 0.059, respectively. The interpopulation overall gene flow was calculated to be 3.99. Thus, the microsatellite loci revealed a moderate genetic differentiation level and constant gene flow between equine populations. However, genetic drift was considered as the main factor of the observed genetic differentiation between the FRI, PRZ, and the majority of the other breeds (Nm < 1.0; Supplementary Table S4 in SOM). The highest drift rate was found between the FRI and PRZ (F_{ST} = 0.370, Nm = 0.43). The high F_{ST} value implies that the genetic variation is explained by the population structure, mainly conditioned by the restricted gene flow between the distinct breeding populations. Then, as expected, the two populations do not share their genetic material with the remaining forty-one populations examined and are isolated from one another. Conversely, only subtle genetic differentiation was established between warmblood populations. Negative F_{ST} values and Nm were recorded in some comparisons (CZW–BAV, CZW–SLW), and these equalled to zero and ∞, respectively. Such a condition roughly indicated that individuals from different populations are genetically closer than those within a population; no genetic subdivision was found between the populations, and thus the two pairs of populations analysed do not differ. Our findings indicated evidence of a strong relationship between the CZW–SLW, BAV–CZW, KWPN–OLD, BAV–KWPN, BAV–OLD, and SLW–BAV based upon the small pairwise F_{ST} leading to genetic admixture occurring as a result of relatively high gene migration between the breeds. Furthermore, the low estimated genetic differentiation (F_{ST} = 0.005) between the NOR and SNOR populations raised in the Czech Republic showed that the heavy horse breeds are genetically close, possibly owing to a common historical origin and high gene migration (Nm = 47, Supplementary Table S4 in SOM). Conversely, the CMB was much better differentiated and scored moderate genetic differentiation values in relation to the NOR and SNOR. The lower pairwise F_{ST} value (0.008) estimates between the CSP and WPB breeds suggests the existence of gene migration (Nm = 31, Supplementary Table S4 in SOM) and the occurrence of allele sharing mainly among the populations sampled in the Czech Republic. As a matter of fact, most stallions eligible for breeding within the CSP are registered in more than one stud book, and they are mostly members of the WPB.

The average F_{ST} values indicate that around 5.9% of the total genetic variation were explained by the breeds' differences, with the remaining 94.1% cor-

https://doi.org/10.17221/2/2018-CJAS

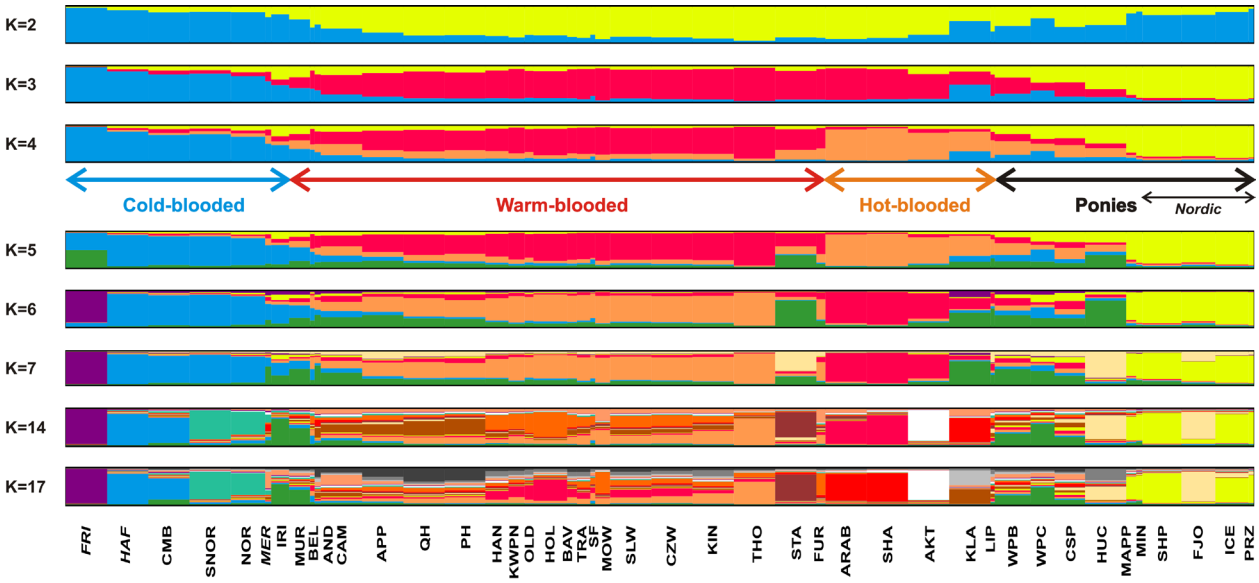

Figure 1. Bayesian model-based clustering for 43 populations (n = 2879)

The estimated population structure is displayed with the population mean Q-scores. The breeds are divided into segments the size and colour of which correspond to the relative proportion of the animal genome assigned to a particular cluster. The highest ΔK was found at K = 4, which indicates the probable number of clusters. The graphical presentations are shown for the last K = 17 because of the high value of H'_{Pop} and ΔK. Breeds not classified in their groups according to the nomenclature are in italics. The breeds abbreviations are as defined in Table 1

responding to the differences between individuals. Such a finding agrees with the overall F_{ST} values described previously in relevant population studies using microsatellite markers (Leroy et al. 2009; Barcaccia et al. 2013; Berber et al. 2014; Gupta et al. 2014; Putnova et al. 2018).

In conclusion, there was no evidence of a serious loss of genetic diversity, and the inbreeding coefficient did not differ significantly from zero. Population differentiation due to genetic structure supported the distinction between the majority of the studied breeds, but a high connectivity and level of breeding between warmbloods kept in the Czech Republic was revealed. Special emphasis was placed on measuring the degree of introgression of the European Warmblood into the Czech Warmblood (with an open stud book) because the admixture pattern and high similarity between Western European horses was expected.

Breed assignment. The overall proportion of individuals correctly assigned to a population was 82.4% as calculated by GENECLASS using a Bayesian method (approach of Rannala and Mountain, 1997). The best individual assignment results (over 90%) were found in 19 populations (Table 1). Within the warmbloods, the worst individual assignment results (below 70%) were yielded by the CZW (32%), SLW (39%), OLD (52%), HAN (64%), KWPN (64%), and BAV (67%). Such values indicate considerable gene migration, close genetic proximity, and no clear distinction between these populations despite their separate stud books. In this study, the CZW and SLW breeds were the most heterogeneous, resulting from the permanent migration of warmblood horses across Europe (e.g. German warmbloods such as the HAN, HOL, OLD, TRA, BAV; the Dutch KWPN; the Belgian BEL; and the French SF). Crossbreeding does not produce stable populations, not even when the world's top gene pool is used for the maximum performance. Assignment success rates below 70% were found also in the PH (68%), CAM (67%), CSP (65%), and WPB (63%). Van de Goor et al. (2011) obtained similar results, namely, that the assignment success is smaller for the Welsh and Warmblood breeds. The outputs demonstrate that an F_{ST} smaller than 0.05 between the majority of Czech warmblood populations is not sufficient to provide the maximum individual assignment success and a near-zero level of "genetic admixture" among the breeds. An individual assignment test revealed a lower assignment success in the NOR (73%) and SNOR (85%) than in the CMB (96%). Additionally, 17% and 13% of individuals from

https://doi.org/10.17221/2/2018-CJAS

the NOR and SNOR populations were assigned in the reverted order. This is in accordance with the fact that SNOR stallions were still used for breeding with NOR mares without major restriction. Their offspring, with more than 50% of the SNOR breed's genes, were thereafter regularly included in the stud book of the SNOR (until 2013) (Vostra-Vydrova et al. 2016).

Genetic structure and Bayesian model-based clustering. In the STRUCTURE analysis (Figure 1), we calculated the highest values of the average symmetric similarity coefficients (H') for the populations (H'_{Pop}) and individuals (H'_{Ind}) depending on the CLUMPP program. The ΔK distribution (Figure 2), along with the different values of the clusters (K = 2–43) for the 43 populations in dependence on Evanno's method, indicated the optimal values of K = 4 (48.7), K = 17 (12.3), and K = 35 (58.5). However, the cluster of K = 35 was situated in the area of the highest standard deviation values for the mean ln of likelihood (lnP(K)); therefore, this cluster was dismissed (Supplementary Figure S1 in SOM). As K increases from 2 to 7, the highest values of H'_{Pop} among the runs (R = 10) are equal to 0.999, 0.998, 0.976, 0.809, 0.863, and 0.846, respectively. The results are shown for the last K = 17 due to the high values of H'_{Pop} (0.864) and ΔK (Figures 1 and 2). The differences between H'_{Pop} and H'_{Ind} were lower than 0.07 in the entire range of K. The STRUCTURE results with the microsatellite data set comprising markers recommended by the FAO (without *HMS1* and X-linked *LEX3*) show no strong differences up to K = 14 (Supplementary Figure S1 in SOM).

Four inferred clusters fit our data set best because of the highest ΔK. For K = 4, there was a clear separation between the cold-blooded horses (also including the FRI, MER, and HAF breeds), hot/warm-blooded ones, and Nordic/pony horses (also including the PRZ). These groups were identified as significant according to the use of the breeds and the morphological characteristics (Leroy et al. 2009; Van de Goor et al. 2011). Some populations (the WPC, WPB, CSP, BEL, AND, CAM, and MUR) were nevertheless contained within multiple groups, but 36 populations were clearly assigned to each cluster (Table 1). As our data suggest, the Czech native breeds including the genetic resources can be also clustered into four groups (the CMB and SNOR ~ cold-blooded; the CSP and HUC ~ pony; the KLA ~ hot-blooded; the MOW, KIN, and CZW ~ warm-blooded; see Table 1). The genetic resources composed of the SNOR, CMB, KLA, and HUC formed the most homogeneous populations from all the Czech native breeds. The CZW, MOW, and KIN clustered together and embodied the most admixed types, together with the CSP. The assignment test isolated the FRI from the other breeds before K = 6, and this breed then maintained its integrity throughout the analysis. The peculiar behaviour of the FRI is probably derived from the high genetic similarity between the individuals, which resulted in the early recognition of the FRI as a separate group during the clustering procedure (Leroy et al. 2009; Van de Goor et al. 2011). The PRZ formed a separate cluster (K = 4) with the FJO, MIN, MAPP, ICE, SHP, and HUC (Table 1). The HUC and FJO left

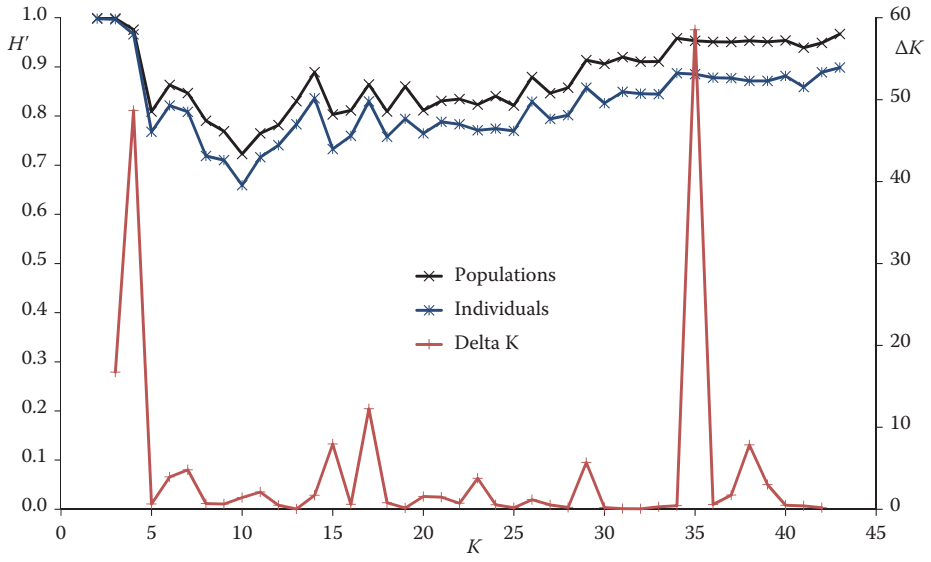

Figure 2. The highest values of symmetric similarity coefficients (H') obtained from 10 runs for populations and individuals, and ΔK distribution along different values of clusters (K = 2–43) for 43 populations

https://doi.org/10.17221/2/2018-CJAS

this formation at K = 6 and K = 10, respectively. All the outgroup PRZ individuals held together with the MIN, MAPP, ICE, and SHP up to K = 29. At K = 30, the assignment isolated the PRZ/ICE and the MIN/MAPP/SHP populations, and these groups did not separate completely until K = 43 (79.6%/65.7% and 43.4%/50.5%/69.5%, respectively). Other populations exhibiting the tendency to hold together until K = 43 were the SNOR/NOR (36%/25.4%) and MOW/FUR (29.8%/24.2%). With the number of clusters fixed at 17–43, the KLA population was split into two subclusters to indicate the existence of subpopulations with the estimated memberships of 0.381 and 0.344 at K = 43, which exactly corresponded to the population subdivision according to the grey and black coat colour varieties. This is not surprising given the results at K =17, where the AND and LIP showed Q-values of assignment to the grey KLA cluster equalling 0.15 and 0.11, respectively (Supplementary Table S6 in SOM). When K became 7, the HUC also formed a separate cluster with the STA (Supplementary Table S5 in SOM), whereas after K had increased to 10, 65.2% of HUC horses combined to form a single cluster with 68.5% of the FJO horses. As K reached 17, most breeds were shared between different clusters (Supplementary Table S6 in SOM). The STA, THO, AKT, FRI, FJO, HUC, KLA, and MER breeds constituted a single cluster, while the QH/PH, SHA/ARAB, and HAF/CMB formed a single cluster at K = 43, 26, and 22, respectively. The HUC at K = 17 (Supplementary Table S6 in SOM) formed its own cluster (31.5%) and also one with the FJO (36.4%), but when K = 25, the HUC (60.5%) and FJO (80.7%) clustered in their own pre-defined populations and were clearly separated from each other. The results are consistent with the breeds' documented history, where the representation of the purebred FJO horses (http://www.hucul-achhk.cz) is tolerated and permitted in the origin of the HUC horse to a certain extent. The Czech WPB population, comprising five sections altogether (the Welsh Mountain Pony, Welsh Pony, Welsh Pony of the Cob type, Welsh Cob, Welsh Part Bred), was analysed separately from the Welsh Pony and Cob stud book. By contrast, the WPB have a minimum of 12.5% registered Welsh blood in their parentages, which can come from the sire, the dam, or both. Indeed, the existence of several subpopulations within the Welsh breed (the Wahlund effect) is expected. The assignment using STRUCTURE (K = 4) classified 39% of the WPB individuals as hot-blooded, while 32% and 36% of the WPC animals were placed among the cold-blooded and pony breeds, respectively (Table 1). In particular, the WPC population was more homogeneous compared to the WPB (Supplementary Table S6 in SOM).

We obtained sufficient intra-breed diversity in the THO kept in the Czech Republic. The inter-breed analysis then showed a clear influence of the THO on many other breeds, such as the BEL, SF, MOW, TRA, BAV, KWPN, KIN, CZW, SLW, QH, and PH; these members were assigned to the THO cluster at 13, 21, 16, 14, 17, 16, 29, 18, 16, 13, and 16%, respectively (K = 17, Supplementary Table S6 in SOM). No notable relationship with the ARAB was observed. The pairwise F_{ST} values between the THO and ARAB (0.0841) did not suggest divergence lower than that observed between the THO and the majority of the other breeds. In addition, at K = 17 the ARAB assigned to the THO cluster at only 1.95% (Supplementary Table S6 in SOM). Although the assumed significant influence of the ARAB on the THO has not been proved to date, there are possible explanations for why the effect is not more apparent. Possibly, the current Czech-based ARAB samples may not reflect the ARAB lineage(s) influential in the founding of the THO. Finally, as noted above and also suggested elsewhere (Bower et al. 2011 – using mitochondrial DNA; Petersen et al. 2013 – using whole-genome single-nucleotide polymorphism data), it may simply be that ARAB bloodlines were not as instrumental in the THO as once thought or that the initial ARAB influence (and genes) was selected against or lost to drift during the development of the modern THO racehorse (Petersen et al. 2013).

CONCLUSION

The present article constitutes the most comprehensive in-depth study mapping the genetic population structure and the degree of admixture within and between worldwide horse breeds kept in the Czech Republic. The individual assignment was performed using the information from 17 microsatellite markers genotyped on 2879 individuals. The analysed outgroup consisted in *Equus przewalskii*. A posterior Bayesian approach implemented in the STRUCTURE program revealed a hierarchical

https://doi.org/10.17221/2/2018-CJAS

dynamic genetic structure with four clusters. The cluster analysis provided an accurate representation of the current genetic relationships between the breeds. While most of the populations were genetically distinct from each other and well-arranged with solid breed structures, some of the entire set showed signs of admixture and/or fragmentation. The comprehensive information about the population stratification and individual assignment success rates can be useful for the development of breeding strategies and conservation programs.

Acknowledgement. The authors are due to the Centre for Research and Utilization of Renewable Energy where the research was carried out. The authors also acknowledge prof. Petr Hořín (Department of Animal Genetics, VFU Brno) for providing the samples of the Camargue, Murgese, and Icelandic horses.

REFERENCES

Barcaccia G., Felicetti M., Galla G., Capomaccio S., Cappelli K., Albertini E., Buttazzoni L., Pieramati C., Silvestrelli M., Supplizi A.V. (2013): Molecular analysis of genetic diversity, population structure and inbreeding level of the Italian Lipizzan horse. Livestock Science, 151, 124–133.

Berber N., Gaouar S., Leroy G., Kdidi S., Tabet Aouel N., Saidi Mehtar N. (2014): Molecular characterization and differentiation of five horse breeds raised in Algeria using polymorphic microsatellite markers. Journal of Animal Breeding and Genetics, 131, 387–394.

Bower M.A., Campana M.G., Whitten M., Edwards C.J., Jones H., Barrett E., Cassidy R., Nisbet R.E.R., Hill E.W., Howe C.J., Binns M. (2011): The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares. Biology Letters, 7, 316–320.

Earl D.A., von Holdt B.M. (2012): STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4, 359–361.

Evanno G., Regnaut S., Goudet J. (2005): Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology, 14, 2611–2620.

Fornal A., Radko A., Piestrzynska-Kajtoch A. (2013): Genetic polymorphism of Hucul horse population based on 17 microsatellite loci. Acta Biochimica Polonica, 60, 761–765.

Goudet J. (2001): FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www.unil.ch/izea/softwares/fstat.html (accessed Dec 24, 2017).

Gupta A.K., Chauhan M., Bhardwaj A., Gupta N., Gupta S.C., Pal Y., Tandon S.N., Vijh R.K. (2014): Comparative genetic diversity analysis among six Indian breeds and English Thoroughbred horses. Livestock Science, 163, 1–11.

Jakobsson M., Rosenberg N.A. (2007): CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23, 1801–1806.

Leroy G., Callede L., Verrier E., Meriaux J.C., Ricard A., Danchin-Burge C., Rognon X. (2009): Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. Genetics Selection Evolution, 41, 5.

Nordborg M., Hu T.T., Ishino Y., Jhaveri J., Toomajian C., Zheng H., Bakker E., Calabrese P., Gladstone J., Goyal R., Jakobsson M., Kim S., Morozov Y., Padhukasahasram B., Plagnol V., Rosenberg N.A., Shah C., Wall J.D., Wang J., Zhao K., Kalbfleisch T., Schulz V., Kreitman M., Bergelson J. (2005): The pattern of polymorphism in Arabidopsis thaliana. PLoS Biology, 3, 1289–1299.

Park S.D.E. (2001): The Excel Microsatellite Toolkit. Trypanotolerance in West African cattle and the population genetic effects of selection. University of Dublin: Ph.D. Thesis.

Petersen J.L., Mickelson J.R., Cothran E.G., Andersson L.S., Axelsson J., Bailey E., Bannasch D., Binns M.M., Borges A.S., Brama P., da Camara Machado A., Distl O., Felicetti M., Fox-Clipsham L., Graves K.T., Guerin G., Haase B., Hasegawa T., Hemmann K., Hill E.W., Leeb T., Lindgren G., Lohi H., Lopes M.S., McGivney B.A., Mikko S., Orr N., Penedo M.C., Piercy R.J., Raekallio M., Rieder S., Roed K.H., Silvestrelli M., Swinburne J., Tozaki T., Vaudin M., Wade C.M., McCue M.E. (2013): Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS ONE, 8, e54997.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L., Estoup A. (2004): GeneClass2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity, 95, 536–539.

Pritchard J.K., Stephens M., Donnelly P. (2000): Inference of population structure using multilocus genotype data. Genetics, 155, 945–959.

Putnova L., Stohl R., Vrtkova I. (2018): Genetic monitoring of horses in the Czech Republic: A large-scale study with a focus on the Czech autochthonous breeds. Journal of Animal Breeding and Genetics, 135, 73–83.

Rannala B., Mountain J.L. (1997): Detecting immigration by using multilocus genotypes. Proceedings of the National

https://doi.org/10.17221/2/2018-CJAS

Academy of Sciences of the United States of America, 94, 9197–9201.

Rosenberg N.A. (2004): Distruct: A program for the graphical display of population structure. Molecular Ecology Notes, 4, 137–138.

Rousset F. (2008): Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources, 8, 103–106.

Slatkin M., Barton N.H. (1989): A comparison of three indirect methods for estimating average levels of gene flow. Evolution, 43, 1349–1368.

Van de Goor L.H., van Haeringen W.A., Lenstra J.A. (2011): Population studies of 17 equine STR for forensic and phylogenetic analysis. Animal Genetics, 42, 627–633.

Vostra-Vydrova H., Vostry L., Hofmanova B., Krupa E., Vesela Z., Schmidova J. (2016): Genetic diversity within and gene flow between three draught horse breeds using genealogical information. Czech Journal of Animal Science, 61, 462–472.

Received: 2017–12–24
Accepted: 2018–09–19