ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/j.1365-2052.2010.02067.x

Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers

Y. H. Ling*,†, Y. H. Ma*,†, W. J. Guan*,†, Y. J. Cheng*,†, Y. P. Wang*,†, J. L. Han*,†, L. Mang‡, Q. J. Zhao*,†, X. H. He*,†, Y. B. Pu*,† and B. L. Fu*,†

*Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China. †The Key Laboratory for Farm Animal Genetic Resources and Utilization of Ministry of Agriculture of China, Beijing 100193, China. ‡College of Animal Science and Veterinary Medicine, Inner Mongolia Agricultural University, Huhhot 010018, China

Summary

We determined the genetic diversity and evolutionary relationships among 26 Chinese indigenous horse breeds and two introduced horse breeds by genotyping these animals for 27 microsatellite loci. The 26 Chinese horse breeds come from 12 different provinces. Two introduced horse breeds were the Mongolia B Horse from Mongolia and the Thoroughbred Horse from the UK. A total of 330 alleles were detected, and the expected heterozygosity ranged from 0.719 (Elenchuns) to 0.780 (Dali). The mean number of alleles among the horse breeds ranged from 6.74 (Hequ) to 8.81 (Debao). Although there were abundant genetic variations found, the genetic differentiation was low between the Chinese horses, which displayed only 2.4% of the total genetic variance among the different breeds. However, genetic differentiation (pairwise *FST*) among Chinese horses, although moderate, was still apparent and varied from 0.001 for the Guizou–Luoping pair to 0.064 for the Jingjiang–Elenchuns pair. The genetic differentiation patterns and genetic relationships among Chinese horse breeds were also consistent with their geographical distribution. The Thoroughbred and Mongolia B breeds could be discerned as two distinct breeds, but the Mongolia B horse in particular suffered genetic admixture with Chinese horses. The Chinese breeds could be divided into five major groups, i.e. the south or along the Yangtze river group (Bose, Debao, Wenshan, Lichuan, Jianchang, Guizhou, Luoping, Jinjiang and Dali), the Qinghai-Tibet Plateau group (Chaidamu, Hequ, Datong, Yushu, Tibet Grassland and Tibet Valley), the Northeast of China group (Elenchuns, Jilin and Heihe), the Northwest of China group (Kazakh, Yili and Yanqi) and the Inner Mongolia group (Mongolia A, Sanhe, Xinihe,Wuzhumuqin and Sengeng). This grouping pattern was further supported by principal component analysis and structure analysis.

Keywords admixture, Chinese indigenous horse, genetic diversity, genetic relationship, microsatellite.

Introduction

Horses were domesticated from the widespread matrilineal integration and taming of wild horses approximately 6000 years ago (Vilà *et al.* 2001). Today, over 300 horse breeds are recognized worldwide, a great number of which are indigenous to China (FAO 2007). China has a long history of horse breeding, and various indigenous breeds have been developed by selection under different ecological conditions. These may be roughly grouped into three main categories in accordance with their geographical region i.e. North Pastoral Grassland, Northwest Plateau and South-west Mountain Area.

Chinese horses serve many functions: the Mongolian Horse is capable of racing high speeds over short distance, and is thus primarily used for riding and carting; the Kazak Horse has a good milking performance; the Wuzhumuqin Horse is used for both riding and drafting, is adapted to hard conditions and is a hard working animal; the Hequ Horse is suitable for working as a cart horse and has a draught force equivalent to 80% of its body weight; and the Yushu Horse is adapted to a plateau climate and can walk freely in swamps, over steep slopes and along narrow winding roads (Xie 1987). Many horses distributed in the Yunnan,

Address for correspondence

Y. H. Ma, Institute of Animal Science, Chinese Academy of Agricultural Sciences, No. 2 West Yuanmingyuan Road, Beijing 100193, China.
E-mail: yuehui.ma@263.net

Accepted for publication 11 March 2010

Guizhou, Sichuan and Guangxi provinces of China, such as the Baise, Debao and Jianchang, do not exceed a height of 1 m at the withers, even as adults.

An understanding of the evolutionary history of domestic breeds and data on genetic variation within and among breeds is vital to the initiatives to provide critically important data for the conservation decision-making process (Rege & Gibson 2003). Little immediate threat to the genetic diversity of horses has been thus far apparent, but all horse breeds will likely experience declines in their overall population numbers as a result of increased mechanization of agriculture and changing human lifestyles. Indeed, some evidence of this threat has already become apparent in certain breeds. To date, most studies of genetic diversity in horses (Aberle *et al.* 2004; Glowatzki *et al.* 2005; Marletta *et al.* 2006) and other livestock species have been carried out on local geographical (national) scales (Saitbekova *et al.* 1999; Martinez *et al.* 2000; Arranz *et al.* 2001; Mateus *et al.* 2004). While such studies are essential for regional breed-specific management and conservation programmes, it is also important to assess how genetic diversity is partitioned on larger geographical scales to better implement region-specific conservation measures (Bruford *et al.* 2003). In our present study, we employed a panel of 27 autosomal microsatellite markers to evaluate how genetic diversity is partitioned within and among a diverse sample of 26 Chinese horse breeds and two introduced breeds. We also considered the impact of geographical location and breed type in determining diversity and differentiation, and examined the extent of admixture among breeds in relation to conservation and management.

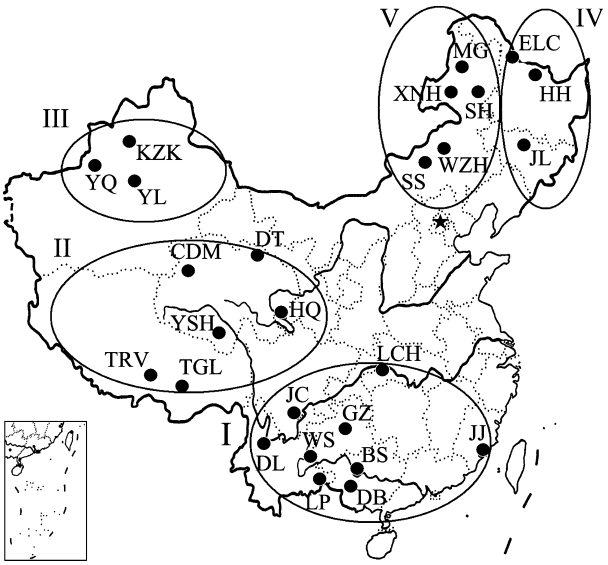

Figure 1 Geographical location of 26 Chinese domestic horse breeds sampled in this study. Jinjiang (JJ) - Jinjiang County, Fujian; Debao (DB) - Debao County, Guangxi; Bose (BS) - Baise County, Guangxi; Guizhou (GZ) - Anshun City, Guizhou; Luoping (LP) - Louping County, Yunnan; Wenshan (WS) - Wenshan County, Yunnan; Dali (DL) - Jianchuan County, Yunnan; Lichuan (LCH) - Lichuan City, Hubei; Jianchang (JC) - Butuo County, Sichuan; Tibet Valley (TRV) - Jiangzi County, Tibet; Tibet Grassland (TGL) - Langkazi County, Tibet; Yushu (YSH) - Yushu City, Qinghai; Hequ (HQ) - Henan County, Qinghai; Chaidamu (CDM) - Dulan County, Qinghai; Datong (DT) - Qilian County, Qinghai; Yili (YL) - Yili County, Xinjiang; Yanqi (YQ) - Yanqi County, Xinjiang; Kazakh (KZK) - Xinyuan County, Xinjiang; Jilin (JL) - Baicheng City, Jilin, Elenchuns (ELC) - Heihe City, Heilongjiang; Heihe (HH) - Sunwu County Heilongjiang; Sanhe (SH) - Hulunbeie City, Inner Mongolia; Sengseng (SS) - Xilinguole City, Inner Mongolia; Wuzhumuqin (WZH) - Xilinguole City, Inner Mongolia; Xinihe (XNH) - Hulunbeie City, Inner Mongolia; Mongolia A (MG) - Hulunbeie City, Inner Mongolia. I: Yangtze river group; II: Qinghai-Tibet Plateau group; III: Northwest group; IV: Northeast group; V: Inner Mongolia group.

Materials and methods

Sample collection and DNA extraction

Blood samples from 1273 animals were collected from 26 domesticated horse breeds from China and two introduced horse breeds (Mongolia B Horse and Thoroughbred Horse). The sample set did not include any sets of parents and offspring. The Thoroughbred Horse was used as outgroup in constructing the dendrogram tree. The 26 Chinese horse breeds come from 12 different provinces (Fig. 1), and our aim was to collect at least 30 samples from a minimum of two separate flocks, although this was not possible for all breeds (more information about these breeds is showed in Table S1). Genomic DNA was recovered from all blood samples using a standard phenol-chloroform extraction method (Sambrook *et al.* 1989).

Microsatellite loci and genotyping

All 27 microsatellite loci recommended by the Food and Agricultural Organization (FAO) and the International Society for Animal Genetics (ISAG) for evaluating horse genetic diversity were analysed (Hoffmann *et al.* 2004). Forward primers were 5′-labelled with fluorescent dyes (FAM or HEX, Sangon Biotech (Shanghai) Co., Ltd.). The relative information specific to these loci is shown in Table S2. Genotypes for each marker were determined using an ABI 3130 DNA Sequencer (Applied Biosystems) with GENESCAN™ – LIZI 500 internal size standards (Applied Biosystems).

PCR amplifications were performed in 12 μl reaction volumes containing 50 ng of genomic DNA, 2.5 mM $MgCl_2$, 250 μM of each dNTP, 0.025 μM of each primer, 1.25 units of *Taq* polymerase and 1 × Magnesium-free PCR buffer (Takara, Japan). Amplifications were carried out using the GENEAMP PCR 9700 thermocycler (Applied Biosystems), with the following cycling parameters: 94 °C for 5 min followed by 30 cycles of 94 °C for 30 s, annealing at 55 °C∼60 °C for 30 s, 72 °C for 30 s and a final step at 72 °C for 10 min. PCR products were diluted to 1/2–1/4 concentrations and 0.75 μl of each diluted product was then mixed with

an internal standard (GENESCAN™-500 LIZ™, Applied Biosystems) according to the manufacturer's instructions. Genotyping was carried out on an ABI 3130xl automated capillary sequencer. Fragment analysis was performed using GENEMAPPER V3.7 software (Applied Biosystems). The third-order least squares method was used for allele size determination (Mburu *et al.* 2003). Genotyping was repeated once if individual samples failed to amplify.

Statistical analysis

For each horse breed, the following statistics were calculated using POPGENE 1.31 (http://www.ualberta.ca/~fyeh): total number of alleles (TNA), number of effective alleles (NEA) and expected heterozygosity (He). The private alleles were counted using the GDA program (http://lewis.eeb.uconn.edu/lewishome/). The mean number of alleles (MNA), observed heterozygosity (Ho), Nei's unbiased gene diversity ($\hat{H}$) (Nei 1987) and standard deviations or variances per breed were computed using the EXCEL MICROSATELLITE TOOLKIT Version 3.1 (http://oscar.gen.tcd.ie/sdepark/ms-toolkit). The F-statistic values (FIT, FIS and FST; Weir & Cockerham 1984), together with the total number of alleles per locus and the allelic richness (AR) of each breed, were estimated with FSTAT 2.9.3 (Goudet 2001). Allelic richness is a measure of the number of alleles that can be corrected to be independent from the sample size (Petit *et al.* 1998). In the present study, allelic richness was calculated using a rarefied sample size of 24 diploid individuals per breed. The *P*-value for FIS within samples was based on 756 000 randomizations. A Fisher's exact test was performed to determine possible deviations from the HWE and genotype linkage disequilibrium using GENEPOP 3.4 (Raymond & Rousset 1995). Exact *P*-values were estimated from the Markov chain algorithm using 10 000 dememorization steps, 500 batches and 5000 iterations per batch. The numbers of loci with significant heterozygote deficit and excess were also obtained.

Nei's standard genetic distance D_S (Nei 1972) and Nei's genetic distance D_A (Nei *et al.* 1983) between breeds were calculated using the DISPAN package. Neighbour-joining (NJ) dendrograms and the unweighted pair group method with arithmetic mean (UPGMA) were constructed based on D_A and D_S. The robustness of the dendrograms was evaluated using a bootstrap test of 1000 resamplings of loci, with replacement.

Principal component analysis (PCA) was performed to reveal major patterns of genetic variability and clustering of breeds based on allele frequencies using MVSP 3.1 (http://www.kovcomp.com). The population genetic structure was revealed using a Bayesian clustering algorithm and STRUCTURE 2.1 software (Pritchard *et al.* 2000). An admixture model with correlated allele frequencies was adopted. The parameter for individual admixture alpha was set to be the same for all clusters with a uniform prior. Posterior probability values for *K* [log likelihood, ln $P(D)$] were estimated by assigning priors from 2 to 28, with five independent runs of each. Structure was run with 10 000 burn-in steps and 100 000 MCMC replicates in all simulations. Because likelihood maximization intrinsically favours partitions with more clusters and the ln$P(D)$ did not provide an unequivocal number of clusters, the number of *K* clusters was also selected based on the log likelihood ratio test, using the formula $-2([\ln P(D)_k]-[\ln P(D)_{k-1}])$, with $df = df_k - df_{k-1}$ (Crawley 1993).

Finally, the graphical displays of population structure were generated using the DISTRUCT program (Rosenberg 2004; http://rosenberglab.bioinformatics.med.umich.edu/distruct.html). These displays represented the fraction of the individual genome that had sub-population ancestry. A synthetic map illustrating geographical variation and admixture coefficients (Q matrix) was generated using Kriging standard deviations methods in GOLDEN SOFTWARE SURFER 8.0 (Golden software Inc. http://www.goldensoftware.com).

Results

Polymorphism of microsatellite loci and alleles

The HWE was tested for all breed-locus combinations. Significant ($P < 0.05$) deviations from a HWE were observed for 102 (13.5%) of 756 breed-locus combinations (Table S3). However, no significant ($P > 0.05$) deviation from a HWE was detected for either a single locus across all breeds or a single breed across all loci. However, on average, 3.6 loci per breed and 3.8 breeds per locus deviated significantly from HWE. The Chaidamu Horse showed the maximum number of loci in disequilibrium (13 loci), followed by Hequ (10 loci).

There was a high allelic variation found in the Chinese horse breeds for 27 loci. A total of 330 alleles were detected, and the mean number of alleles across the 27 loci was 12.2. The number of alleles per locus (*At*) ranged from 7 for *LEX34* to 19 for *ASB17* (mean 12.2), which is greater than that reported by FAO–ISAG (Table 1). Allelic richness over all samples per locus (*Rt*) was measured at between 4.6 (HTG07) and 13.6 (ASB17) and the mean allelic richness across 27 loci in our Chinese horse breeds was 7.8. Generally, these loci were polymorphic, and the polymorphism information content (PIC) across the 27 loci ranged between 0.500 (*HMS45*) and 0.851 (*COR058*).

Genetic diversity within breeds

A summary of the identified polymorphisms from all of the horse breeds we tested is listed in Table 2. For the Chinese horses, the observed heterozygosity (Ho) ranged from 0.691 to 0.770 (mean value: 0.743; SD: 0.013), with the lowest value found in the Elenchuns and the highest in the Kazakh breed. The outgroups, Mongolia B and Thoroughbred, were 0.728 (SD = 0.014) and 0.673 (SD = 0.016) respectively.

Table 1 Basic genetic parameters and *F*-statistics for the 26 Chinese horse breeds at 27 microsatellite loci.

Locus	*At*	*Rt*	*PCI*	*FIT*	*FST*	*FIS*
HMS06	9	6.3	0.721	0.036**	0.028***	0.008
HMS07	10	7.4	0.722	0.038**	0.012***	0.026*
HTG07	10	4.6	0.654	0.036	0.027***	0.009
HMS03	14	8.0	0.765	0.075***	0.024***	0.052***
COR082	14	7.7	0.764	0.057***	0.025***	0.032***
COR069	11	6.9	0.728	0.045**	0.026***	0.020
HTG04 (AF169165)	8	5.9	0.575	0.031*	0.022***	0.009
LEX54 (AF075656)	13	7.6	0.706	0.028*	0.024***	0.004***
COR058	18	11.0	0.851	0.024*	0.017***	0.007
ASB17	19	13.6	0.840	0.061***	0.022***	0.040
SGCV28	9	5.8	0.682	0.033*	0.039***	−0.006
HMS02 (X74631)	13	8.5	0.786	0.039**	0.026***	0.014
COR022	8	5.5	0.668	0.017	0.018***	−0.001
HMS45	8	5.7	0.500	−0.006	0.031***	−0.039
LEX34 (AF075636)	7	5.4	0.559	0.021	0.025***	−0.004
COR071	13	7.6	0.740	0.073***	0.020***	0.054***
LEX73 (AF213359)	17	11.0	0.797	0.071***	0.033***	0.038**
HTG10	13	9.0	0.756	0.100***	0.023***	0.079***
COR007	13	8.0	0.761	0.020	0.021***	−0.001
LEX63 (AF075663)	12	8.4	0.685	0.133***	0.028***	0.108***
ASB23	16	9.1	0.786	0.048***	0.025***	0.023*
COR018	13	6.4	0.727	0.041**	0.023***	0.019
VHL20	13	9.0	0.809	0.047***	0.022***	0.026*
AHT04 (Y07733)	15	9.3	0.809	0.024*	0.017***	0.007
UCDEQ425	11	8.4	0.716	0.054**	0.039***	0.015
HTG06	11	6.6	0.562	−0.032	0.017***	−0.049
ASB02	12	8.4	0.786	0.081***	0.019***	0.063***
Total	330			0.046***	0.024***	0.023***

GenBank accession numbers are in parentheses for those markers where searching using locus name does not bring up sequence.

At, total number of alleles per locus; Rt, allelic richness over all samples; PIC, polymorphic information content; FIT, fixation indices of total population; FST, fixation index resulting from comparing subpopulations to the total population; FIS, fixation indices of subpopulation. Significant levels of deficit in heterozygotes, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

There were notable differences in the $\hat{H}$ and *AR* among these breeds. The unbiased gene diversity (Ĥ) values varied between 0.719 in Elenchuns horses and 0.780 in Dali horses (mean value: 0.760; SD: 0.018). Allelic richness was the highest in the Dali breed (AR = 8.24) and lowest in Elenchuns (AR = 6.61), with a mean of 7.21. Allelic richness in the Mongolia B and Thoroughbred horses was 7.61 and 4.65 respectively.

For Chinese horse breeds, the MNA among our horse breeds ranges from 6.74 (Hequ) to 8.81 (Debao). The number of effective alleles showed a pattern among these breeds that was similar to that of allelic richness. The mean Ĥ and AR values for Mongolia B (WMG) were similar to the Inner Mongolia horses (MG). The Thoroughbred horses had significantly lower Ĥ and AR levels than the other breeds analysed in this study. A total of 63 private alleles (NPA) were amplified in our cohort, 60 unique to Chinese horses and three unique to Mongolian horses. The NPA of the Dali Horse was particularly high (NPA = 14), representing 22% of the total NPA. Most of the private alleles (56) were at very low frequencies of below 5%. Six alleles unique to Chinese horses and one to a Mongolian horse showed a frequency that exceeded 5% (data not shown). These included the Dali horse, which had several private alleles of a frequency equal to or higher than 5%, including 156 bp for *HMS06* (frequency, 5%), 134 bp for *HTG07* (frequency, 5%), 214 bp for *COR082* (frequency, 6.67%), 90 bp for *ASB17* (frequency, 6.67%), and 142 bp for *COR022* (frequency, 8.33%). In addition, the Heihe horse had a high frequency private allele of 165 bp for *AHT04* (also known as HMB4; frequency, 5.59%), and Mongolia B horse of 205 bp for *COR058* (frequency, 6.25%).

Genetic distances and clustering

A neighbour-joining tree was constructed on the basis of the D_A genetic distances with relatively high bootstrap values (Fig. 2a). As expected, the Thoroughbred Horse was most distinct for other breeds first, and the Mongolia B (WMG) was divided from Chinese horse breeds. Mongolia A (MG)

Table 2 Sampling information and basic parameters for the genetic diversity associated with the 28 horse breeds analysed in this study.

				Allelic Diversity					Genetic Diversity		
Area	Breeds	Code	*n*	TNA	MNA	NEA	NPA	AR	$\hat{H}$ (SD)	Ho(SD)	*FIS*
Yangtze river of China	Jinjiang	JJ	60	215	7.96	4.09	2	6.87	0.741(0.016)	0.733(0.011)	0.011
	Debao	DB	66	238	8.81	4.77	5	7.65	0.764(0.020)	0.742(0.010)	0.029**
	Bose	BS	60	225	8.33	4.70	2	7.46	0.762(0.021)	0.741(0.011)	0.028*
	Guizhou	GZ	38	209	7.74	4.62	1	7.21	0.762(0.022)	0.762(0.013)	0
	Luoping	LP	37	208	7.70	4.75	2	7.25	0.771(0.020)	0.744(0.014)	0.036*
	Wenshan	WS	36	214	7.93	4.55	2	7.40	0.770(0.016)	0.754(0.014)	0.021
	Dali	DL	30	231	8.56	4.95	14	8.24	0.780(0.021)	0.729(0.016)	0.069***
	Lichuan	LCH	38	212	7.85	4.75	3	7.26	0.769(0.024)	0.752(0.013)	0.021
	Jianchang	JC	52	209	7.74	4.71	1	7.11	0.775(0.016)	0.740(0.012)	0.045***
Qinghai-Tibet plateau of China	Tibet Valley	TRV	52	215	7.96	4.56	2	7.33	0.765(0.016)	0.748(0.012)	0.022*
	Tibet Grassland	TGL	61	225	8.33	4.33	1	7.36	0.755(0.015)	0.746(0.011)	0.012
	Yushu	YSH	50	227	8.41	4.57	1	7.58	0.764(0.016)	0.724(0.012)	0.052***
	Hequ	HQ	24	182	6.74	4.41	0	6.74	0.749(0.024)	0.752(0.017)	–0.004
	Chaidamu	CDM	33	202	7.48	4.72	2	7.21	0.769(0.019)	0.741(0.015)	0.037*
	Datong	DT	56	227	8.41	4.61	2	7.57	0.763(0.018)	0.737(0.011)	0.035**
Northwest of China	Yili	YL	60	230	8.52	4.71	1	7.55	0.772(0.015)	0.762(0.011)	0.013
	Yanqi	YQ	55	227	8.41	4.65	2	7.50	0.767(0.015)	0.747(0.011)	0.027*
	Kazakh	KZK	50	228	8.44	4.92	2	7.65	0.773(0.018)	0.770(0.011)	0.005
Northeast of China	Jilin	JL	52	212	7.85	4.39	4	7.02	0.752(0.019)	0.741(0.012)	0.014
	Elenchuns	ELC	30	186	6.89	3.87	1	6.61	0.719(0.022)	0.691(0.016)	0.039*
	Heihe	HH	42	217	8.04	4.54	3	7.35	0.763(0.016)	0.758(0.013)	0.006
Inner Mongolia of China	Sanhe	SH	31	215	7.96	4.54	0	7.60	0.768(0.015)	0.744(0.015)	0.031*
	Sengseng	SS	40	198	7.33	4.12	1	6.81	0.740(0.018)	0.727(0.014)	0.016
	Wuzhumuqin	WZH	50	216	8.00	4.63	2	7.27	0.758(0.018)	0.751(0.012)	0.009
	Xinihe	XNH	50	215	7.96	4.39	4	7.18	0.760(0.014)	0.735(0.012)	0.033**
	Mongolia A	MG	50	205	7.59	4.28	0	6.82	0.740(0.020)	0.747(0.012)	–0.010
	Mean							7.29	0.760(0.018)	0.743(0.013)	0
Mongolia	Mongolia B	WMG	40	224	8.30	4.54	3	7.61	0.768(0.015)	0.728(0.014)	0.053***
Britain	Thoroughbred	THB	30	129	4.78	3.16	0	4.65	0.646(0.029)	0.673(0.016)	–0.042
	All Mean							7.21	0.757(0.018)	0.740(0.013)	

n, sample size; TNA, total number of alleles; NEA, number of effective alleles; MNA, mean number of alleles; AR, allelic richness; NPA, number of private alleles; $\hat{H}$, unbiased gene diversity; Ho, observed heterozygosity; SD, standard deviation; HWE, number of loci that deviate from a Hardy–Weinberg equilibrium; FIS, population inbreeding coefficient.
$^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

and Mongolia B (WMG) are subpopulations of the Mongolia Horse that come from different countries, but they were not classed into the same cluster. This is possibly because they were each intercrossed with other horse breeds in different districts.

Among all horse breeds, the Chinese breeds clustered closely together. Discarding the Thoroughbred Horse and the Mongolia B Horse, we obtained a new NJ tree with five main groups (Fig. 2b), which generally corresponded to five geographical regions, i.e. the south or along the Yangtze river group (I: Bose, Debao, Wenshan, Lichuan, Jianchang, Guizhou, Luoping, Jinjiang and Dali), the Qinghai-Tibet Plateau group (II: Chaidamu, Hequ, Datong, Yushu, Tibet Grassland and Tibet Valley), the Northwest of China group (III: Kazakh, Yili and Yanqi), the Northeast of China group (IV: Elenchuns, Jilin and Heihe) and Inner Mongolia group (V: Mongolia A, Sanhe, Xinihe,Wuzhumuqin and Sengeng). The Northeast of China group (Elenchuns, Jilin and Heihe) has a low boostrap value, which might be an indication that these breeds have been hybridized more frequently, losing their genetic distinctiveness.

This grouping pattern was further supported by PCA analysis. Since phylogenetic reconstruction may not easily take into account the effects of admixture between breeds, we performed the PCA method to further investigate possible genetic relationships between Chinese horse breeds. The first principal component (PC) explains 23% of the observed genetic variation, and the second and third PCs resolve 14 and 10% of this variation respectively. Together, therefore, these three PCs account for 47% of the total

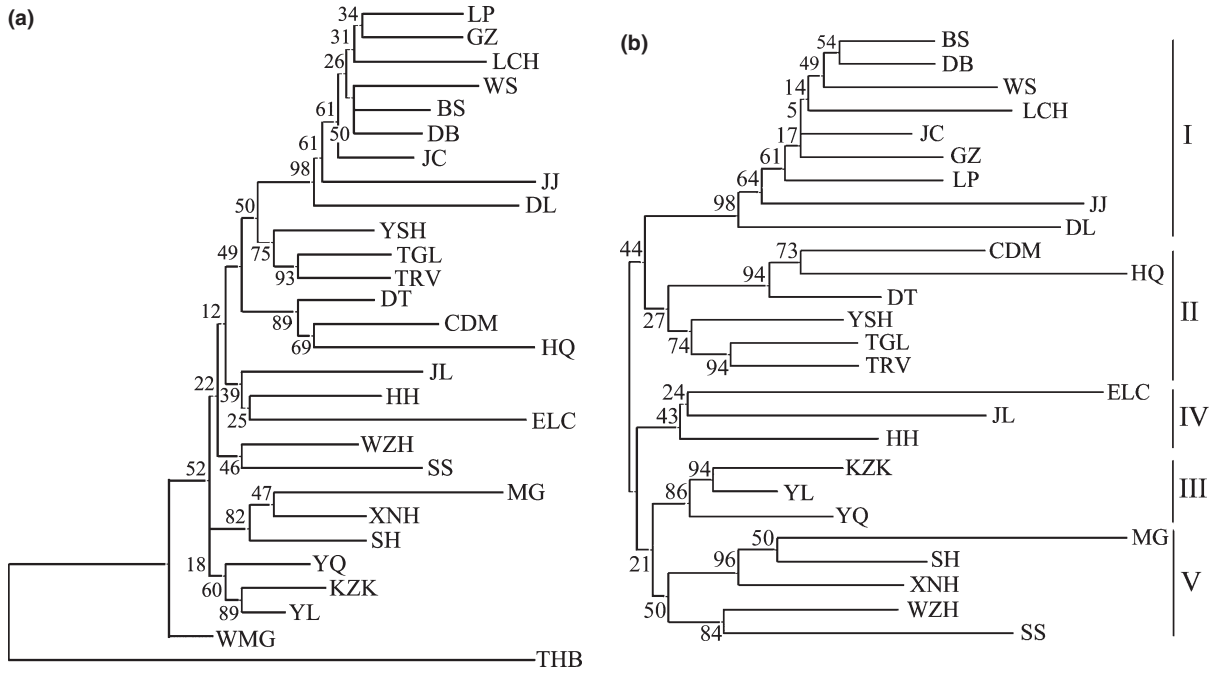

Figure 2 Neighbour-joining (NJ) trees for (a) 26 Chinese indigenous horse breeds and two introduced horse breeds and (b) the 26 Chinese indigenous horse breeds only. Abbreviations are as indicated in the legend to Figure 1.

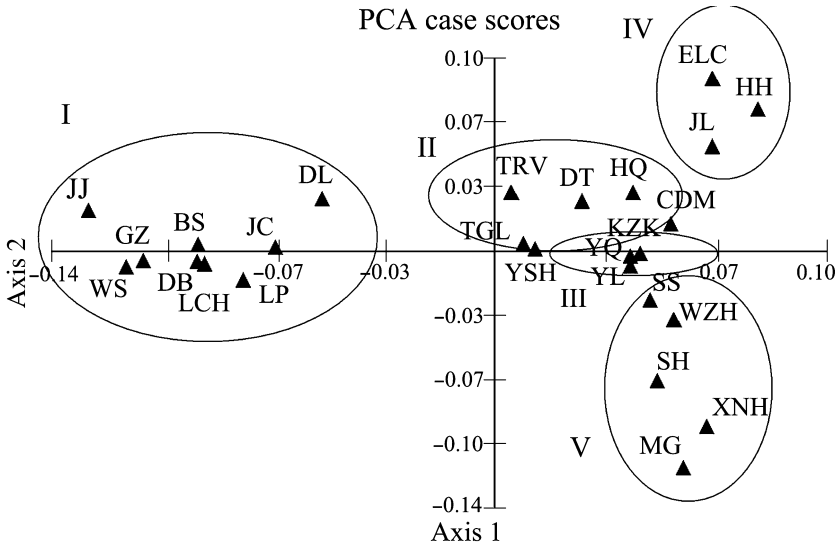

Figure 3 Principal component analysis of 26 Chinese indigenous horse breeds. Axis 1 and Axis 2 represent the first two principal factors, respectively. Abbreviations are as indicated for Table 2.

genetic variation. A two-dimensional scatter plot for the Chinese horse breeds we analysed (Fig. 3) reveals that they are clustered into several different groups and reflects the geographical distribution of these breeds, with the first and the second components corresponding to distributions from north to south and from east to west respectively.

Genetic differentiation and population structure

The inbreeding coefficients (FIS) for 23 of the Chinese horse breeds and the Mongolia B breed (Table 2) were positive, and the values of three Chinese breeds (Dali, Jianchang and Yushu; *FIS* > 0.04) differed significantly from zero ($P < 0.001$). These results indicate a predominance of mating between close relatives. The highest FIS values were found in Dali (0.069) and Yushu (0.052), which was in accordance with the discrepancy between the unbiased gene diversity and observed heterozygosity in these breeds. An excess of heterozygotes (negative FIS value) was observed in two Chinese breeds (Hequ and Mongolia A) and the Thoroughbred breed. Heterozygote excess in these breeds may be attributable to recent admixture. Genetic differentiation among the breeds was characterized by estimating overall and pairwise FST values using FSTAT. The significance levels

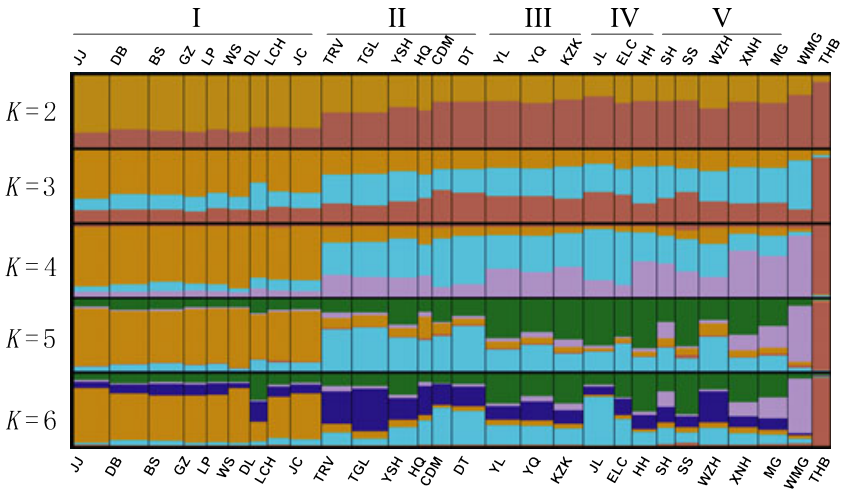

Figure 4 Structure analysis of 26 Chinese indigenous horse breeds and two introduced horse breeds assuming K = 2, 3, 4, 5 and 6. Abbreviations are as indicated in the legend to Figure 1.

for the overall and pairwise FST values were determined after 10 000 permutations. The total FST (0.024) for Chinese horse breeds suggests that 97.6% of the total genetic variation was resulting from genetic differentiation within each breed and thus that only 2.4% existed among breeds.

Genetic differentiation (pairwise FST) among Chinese horses, although moderate, was still apparent, with values ranging from 0.001 for the Guizou–Luoping pair to 0.064 for the Jingjiang–Elenchuns pair (Table S4). The pairwise FST coefficients between Chinese horses and Thoroughbred Horse were significantly high, ranging from 0.128 (between Yili and Thoroughbred) to 0.157 (Jingjiang – Thoroughbred).

Bayesian structure analysis did not return an unequivocal number of genetic clusters i.e. the ln $P(D)$ seemed to level out at seven clusters and then started to fall (Fig. S1). The probability reached a maximum around $K = 6–7$ and showed a sharp fall for $K > 7$. There seemed to be an optimum of six clusters, which was also indicated by a log likelihood ratio test. Based upon these findings, we decided to set $K = 6$ and for K-values greater than this, and variations between runs increased, as revealed by a dramatic drop in similarity coefficients, depicting the absence of additional sub-structuring (Crawley 1993). Figure 4 shows the structure of our 26 Chinese indigenous horse breeds cohort and two introduced horse breeds, assuming $K = 2$, 3, 4, 5 and 6. Subsequently, $K = 6$ was taken to represent most relevant number of genetic clusters in the dataset, because, at this K-value, the Chinese horse breeds were identified as five distinct genetic clusters that corresponded to their breed designations. The horses in the south or along the Yangtze river group were also found to be a distinct and almost genetically homogeneous breed.

Correlations between geographical distances and genetic relationships were significant ($R^2 = 0.214$, $P < 0.0001$). However, the north horse breeds of China (Elenchuns, Jilin, Heihe, Kazakh, Yili, Yanqi, Mongolia A, Sanhe, Xinihe, Wuzhumuqin and Sengseng) showed very different genetic distances among themselves, although the geographical distances were similar to other breeds. If we removed the north horse breeds of China from our calculations, a significant correlation ($R^2 = 0.764$, $P < 0.0001$) emerged for the other breed pairs. Moreover, according to the results of Bayesian structure analysis, a synthetic map (Fig. S2) displayed geographical variation and admixture coefficients (Q matrix) among Chinese horse breeds.

Discussion

The level of AR, the number of alleles sampled and heterozygosity found in the Chinese horses analysed in this study were similar to those previously found in Inner Mongolia horses (Li *et al.* 2005) and the Hispano–Breton horses (Pérez-Gutiérrez *et al.* 2008). In general, our current data reveal a higher diversity level in comparison with the previous reports of Spanish horses (Cañon *et al.* 2000; Solis *et al.* 2005; Marletta *et al.* 2006), German draught horses (Aberle *et al.* 2004), Swiss horses (Glowatzki *et al.* 2005), Norwegian horses (Bjørnstad *et al.* 2000), Portuguese Sorraia and Friesian horse breeds (Luís *et al.* 2007) and Japanese native horse breeds (Kakoi *et al.* 2007).

There were a wide range of values concerning heterozygosity and number of rare alleles among Chinese horse breeds. The Dali Horse had particularly high NPA and $\hat{H}$ values. Furthermore, there were 21 Chinese horse breeds which had fewer than three rare alleles. Pure breeding and a high degree of inbreeding have long been practised to emphasize or even set specific traits among a small number of different strains in horses (Forbis 1980), and this has probably led to the loss of rare alleles and a reduction of heterozygosity. In general, direct comparisons with other studies have to be carefully considered, because different microsatellite markers and partly different markers were used in these different reports.

The genetic differentiation (FST) parameter suggests an overall differentiation of 2.4% between the Chinese horse breeds we tested. This is comparable with a previously

reported value of 2.8% for the total genetic variance among Chinese domestic buffalo breeds (Zhang *et al.* 2007) and a 3.6% differentiation observed among Spanish donkey breeds (Aranguren-Méndez *et al.* 2001). Genetic differentiation among three Pantaneiro horse breeds (*F*ST = 0.008–0.064) was also low (Giacomoni *et al.*2008). However, our *FST* value is slightly smaller than that found by some other studies, such as the 7.8% reported by Cañon *et al.* (2000) for seven Spanish horse breeds, 10% reported by Zabek *et al.* (2005) for some Polish breeds, 10.9% reported by Díaz *et al.* (2002) for the Argentine Creole and Thoroughbreds and 9.9% reported by Leroy *et al.* (2009) for French horses.

It has been pointed out by some researchers that the typical high within-breed variability of microsatellites may result in low differentiation values (Hedrick 1999; Balloux & Lugon-Moulin 2002). Thus, the order of magnitude of genetic differentiation between breeds assessed by *FST* estimators seems to be always low and quite constant regardless of the species (MacHugh *et al.* 1998; Arranz *et al.* 2001; Laval *et al.* 2002).

Gene flows between breeds from the Chinese Mongolia Horse (Mongolia A, Wuzhumuqin, Sengseng and Xinihe) and the Mongolia Horse (Mongolia B) have not been possible in recent decades as a result of the division of Mongolia in 1911. The reproductive isolation of these breeds led to significant genetic and genotypic differentiations between the Mongolia Horse subpopulations in China and the Mongolia Horse subpopulation in Mongolia, and thus to the development of a new, genetically distinguishable Mongolian horse breed. Furthermore, the Mongolian breeds in China are in fact not significantly genetically distinguishable from other horse breeds in inner Mongolia. Whilst the Mongolia B subpopulation may be kept as a distinct breed, crossbreeding in this case would offer a good opportunity to increase genetic variability, to decrease inbreeding coefficients, and to stabilize the population size, as these breeds are genealogically of the same origin.

The levels of genetic diversity of the Chinese southern horse breeds we analysed, as assessed by microsatellite methods, were comparable with those reported in other studies of Chinese southern horse breeds using blood protein makers (Hou *et al.* 1993; Wang *et al.* 2001). In the report of Wang *et al.*, polymorphisms were detected at four loci by polyacrylamide gel electrophoresis of serum samples from 16 Chinese miniature horses (Wang *et al.* 2001). These results confirm that Chinese southern horse breeds are highly polymorphic.

When deciding on conservation priorities for animals such as the horse, several factors should be taken into account, including the adaptation to specific environments or diseases and the possession of special traits that have cultural, scientific or future economic value (Ruane 1999). Information regarding both within- and among- breed diversity is therefore very important. The former provides information that is pertinent to management at the population level and the latter helps to identify divergent breeds that may harbour distinct genotypes and are thus worthy of conservation efforts, even if their within-breed diversity is relatively high. Genetic diversity within and between breeds can also influence decisions affecting the breeds or species to be preserved, but choices should be based on objective criteria and computations. It is difficult to base such priorities on subjective criteria such as beauty or interest in future or present generations (Thaon d'Arnoldi *et al.* 1998).

If the aim of conservation is to preserve within-breed genetic diversity. The Dali and Debao breeds, in the case of Chinese horses, possess the most diversity based on TNA, MNA, NEA, NPA and AR, and they are more likely to be able to cope with future challenges. However, if the aim is to preserve between-breed diversity for these animals, the Jinjiang Horse is more distant from the other breeds among the southern horse breeds of China, and its loss would represent the single highest loss of total diversity among these animals. If possible, all populations should be maintained, even those with the smallest contribution to the global diversity, as they may have other value, such as cultural value or utility.

Acknowledgements

This work was supported by the Eleventh Five-year Plan of Country Technology Pedestal (2006BAD13B08) and the Technology Resources Platform of the Ministry of Science and Technology (2005DKA21101). We acknowledge the assistance of many local animal institutes or colleges in China with sample collection.

References

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. *Animal Genetics* **35,** 270–7.

Aranguren-Méndez J., Jordana J. & Gomez M. (2001) Genetic diversity in Spanish donkey breeds using microsatellite DNA markers. *Genetics Selection Evolution* **33,** 433–42.

Arranz J.J., Bayón Y. & San Primitivo F. (2001) Differentiation among Spanish sheep breeds using microsatellites. *Genetics Selection Evolution* **33,** 529–42.

Balloux F. & Lugon-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11,** 155–65.

Bjørnstad G., Gunby E. & Røed K.H. (2000) Genetic structure of Norwegian horse breeds. *Journal of Animal Breeding and Genetics* **117,** 307–17.

Bruford M.W., Bradley D.G. & Luikart G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4,** 900–10.

Cañon J., Checa M.L., Carleos C., Vega-Pla J.L., Vallejo M. & Dunner S. (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Animal Genetics* **31,** 39–48.

Crawley M.J. (1993) *GLIM for Ecologists*. Blackwell Scientific Publications, Oxford, UK.

Díaz S., Dulout F.N. & Peral-García P. (2002) Greater genetic variability in Argentine Creole than in Thoroughbred horses based on serum protein polymorphisms. *Genetics and Molecular Research* **1**, 261–5.

FAO (2007) *The State of the World's Animal Genetic Resources for Food and Agriculture*. Available at http://faostat.fao.org/Desktop Default.aspx?PageID=573&lang=en#ancor (last accessed 7 May 2010).

Forbis J. (1980) *Das klassische arabische Pferd*. Parey, Berlin und Hamburg.

Giacomoni E.H., Fernández-Stolz G.P. & Freitas T.R.O. (2008) Genetic diversity in the Pantaneiro horse breed assessed using microsatellite DNA markers. *Genetics and Molecular Research* **7**, 261–70.

Glowatzki M.L., Muntwyler J., Pfister W., Marti E., Rieder S., Poncet P.A. & Gaillard C. (2005) Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics* **37**, 33–9.

Goudet J. (2001) *FSTAT, A Program to Estimate and Test Gene Diversities and Fixation (Version 2.9.3.2)*. Available at http://www2.unil.ch/popgen/softwares/fstat.htm (last accessed 7 May 2010).

Hedrick P.W. (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–8.

Hoffmann I., Marsan P.A., Barker S.F., Cothran E.G., Hanotte O., Lenstra J.A., Milan D., Weigend S. & Simianer H. (2004) New MoDAD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/FAO working group. In: *29th International Conference on Animal Genetics*, pp. 35–6. FAO, Meiji University, Tokyo, Japan.

Hou W.T., Li X.Y. & Li Q.R. (1993) Studies on the genetic differentiation and relationship some local types in southwestern horses of China. *Acta Agriculturae Boreali-occidentalis Sinica* **4**, 94–8.

Kakoi H., Tozaki T. & Gawahara H. (2007) Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochemical Genetics* **45**, 375–95.

Laval G.M., SanCristobal M. & Chevalet C. (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics Selection Evolution* **34**, 481–507.

Leroy G., Verrier E., Callèle L., Mériaux J.C., Ricard A. & Rognon X. (2009) Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genetics Selection Evolution* **41**, 1–12.

Li J.L., Mang L. & Shi Y.F. (2005) Evaluation of genetic diversity of Mongolian horse and thoroughbred horse using microsatellite markers. *Acta Veterinaria et Zootechnica S inica* **63**, 6–9.

Luís C., Juras R., Oom M.M. & Cothran E.G. (2007) Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Animal Genetics* **38**, 20–7.

MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29**, 333–40.

Marletta D., Tupsc-Yupanqui I., Bordonaro S., García D., Guastella A.M., Criscione A., Cañón J. & Dunner S. (2006) Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellites. *Journal of Animal Breeding and Genetics* **123**, 315–25.

Martinez A.M., Delgado J.V., Rodero A. & Vega-Pla J.L. (2000) Genetic structure of the Iberian pig breed using microsatellites. *Animal Genetics* **31**, 295–301.

Mateus J.C., Penedo M.C.T., Alves V.C., Ramos M. & Rangel-Figueiredo T. (2004) Genetic diversity and differentiation in Portuguese cattle breeds using microsatellites. *Animal Genetics* **35**, 106–13.

Mburu D.N., Ochieng J.W., Kuria S.G., Jianlin H., Kaufmann B., Rege J.E. & Hanotte O. (2003) Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Animal Genetics* **34**, 26–32.

Nei M. (1972) Genetic distances between populations. *The American Naturalist* **106**, 283–92.

Nei M.. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M., Tajima R. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153–70.

Pérez-Gutiérrez L.M., Peña A.D. & Arana P. (2008) Genetic analysis of the Hispano-Breton heavy horse. *Animal Genetics* **39**, 506–14.

Petit R.J., Ei Mousadik A. & Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844–55.

Pritchard J.K., Stephens M. & Donnely P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59.

Raymond M. & Rousset F. (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–9.

Rege J.E.O. & Gibson J.P. (2003) Animal genetic resources and economic development: issues in relation to economic valuation. *Ecological Economics* **45**, 319–30.

Rosenberg N.A. (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137–8.

Ruane J. (1999) A critical review of the values of genetic distance studies in conservation of animal genetic resources. *Journal of Animal Breeding and Genetics* **159**, 699–713.

Saitbekova N., Gaillard C., Obexer-Ruff G. & Dolf G. (1999) Genetic diversity in Swiss goat breeds based on microsatellite analysis. *Animal Genetics* **30**, 36–41.

Sambrook J., Fritsch E.F. & Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Solis A., Jugo B.M., Mériaux J.C., Iriondo M., Mazón L.I., Aguirre A.I., Vicario A. & Estomba A. (2005) Genetic diversity within and among four South European native horse breeds based on microsatellite DNA analysis: implications for conservation. *Journal of Heredity* **96**, 670–8.

Thaon d'Arnoldi C., Foulley J.L. & Ollivier L. (1998) An overview of the Weitzman approach to diversity. *Genetics Selection Evolution* **30**, 149–61.

Vilà C., Leonard J.A., Gotherstom A., Marklund S., Sandberg K., Linden K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474–7.

Wang Z.S., Li J.C. & Guo Y.X. (2001) Preliminary study on genetic polymorphism of transferrin in Chinese miniature horses.

Heilongjiang Journal of Animal Science and Veterinary Medicine **10,** 3–5.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38,** 1358–70.

Xie C.X. (1987) *Horse and Ass Breeds in China*. Shanghai: Shanghai Scientific and Technical Publishing House, Shanghai.

Zabek T., Nogaj A., Radko A., Nogaj J. & Slota E. (2005) Genetic variation of Polish endangered Bilgoraj horses and two common horse breeds in microsatellite loci. *Journal of Applied Genetics* **46,** 299–305.

Zhang Y., Sun D., Yu Y. & Zhang Y. (2007) Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers. *Animal Genetics* **38,** 569–75.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 ln of the probability of the data [ln $P(D)$] and its variance for K from 2 to 10.

Figure S2 Admixture coefficient (Qmatrix) variation spatially among Chinese horse populations.

Table S1 Name, code, sample size and source region of 28 horse breeds.

Table S2 List of 27 microsatellite loci used in the genetic analyses, including annealing temperature (TA), allele size range and forward label.

Table S3 The *P*-value of every population and every locus in Hardy–Weinberg test.

Table S4 Pairwise population differentiation (FST) below the diagonal and geographical distance above the diagonal among 26 Chinese horse breeds.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.