GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic diversity and population structure of Chinese pony breeds using microsatellite markers

L.X. Xu[1], S.L. Yang[1], R.Y. Lin[2], H.B. Yang[1], A.P. Li[1] and Q.S. Wan[3]

[1]Key Laboratory of Animal Genetics,
Breeding and Reproduction for Mountainous Areas,
Chinese Educational Ministry, Guizhou University, Guiyang, China
[2]Key Laboratory of Pig Genetics and Breeding,
Huazhong Agricultural University, Huazhong, China
[3]Animal Husbandry Bureau of Pingbian County, Yunnan, China

Corresponding author: S.L. Yang
E-mail: shenglinyang@126.com

Genet. Mol. Res. 11 (3): 2629-2640 (2012)
Received December 5, 2011
Accepted May 11, 2012
Published June 25, 2012
DOI http://dx.doi.org/10.4238/2012.June.25.4

ABSTRACT. China is one of the principal origins of ponies in the world. We made a comprehensive analysis of genetic diversity and population structure of Chinese ponies based on 174 animals of five indigenous Chinese pony breeds from five provinces using 13 microsatellite markers. One hundred and forty-four alleles were detected; the mean number of effective alleles among the pony breeds ranged from 5.38 (Guizhou) to 6.78 (Sichuan); the expected heterozygosity ranged from 0.82 (Guizhou) to 0.85 (Debao, Sichuan). Although abundant genetic variation was found, the genetic differentiation was low between the ponies, with 6% total genetic variance among the different breeds. All the pairwise F_{ST} values were significant; they varied from 0.0424 for the Sichuan-Yunnan pair to 0.0833 for the Guizhou-Sichuan pair. All five pony breeds deviated from Hardy-Weinberg equilibrium, except the Yunnan pony. Phylogenetic trees of the five pony breeds based on genetic distances were constructed using a neighbor-joining method. The Sichuan and Yunnan ponies were

grouped into the same branch, with a high bootstrap support value (97%). Guizhou and Ningqiang ponies were clustered into the same branch with a bootstrap value of 56%, whereas the Debao pony was placed in a separate group, with a bootstrap value of 56%. This grouping pattern was supported by genetic structure analysis.

Key words: Chinese pony populations; Genetic diversity; Genetic relationship; Microsatellite markers

INTRODUCTION

Horses have played a significant role in China's political, economic and military history during the last century and the past decades. Chinese ponies, as specific and precious companion animals, have had a close relationship with Chinese farmers. Particularly, ponies have been used as important tools for riding, amusement and carting in southwest mountainous areas. Chinese ponies distributed in Yunan, Guizhou, Shanxi, Guangxi and Sichuan Provinces of China do not exceed a height of 106 cm at the withers even as adults (He, 1987). In recent years, policies for the conservation of important genetic resources have been issued by the Chinese government. Although ponies were listed in the precious animal categories for conservation, the number of Chinese ponies has sharply decreased in the past several decades (Wang and Yue, 2002). For instance, the total number of Debao pony in Guangxi Province is less than 300 (Jiang et al., 2011), and the number of ponies in Malipo county of Yunan Province does not exceed 100 (Sun et al., 2007). It is necessary to develop strategies for their conservation.

Up to now, the origin of the Chinese pony remains unclear, with controversy between the theory of independent origin and the theory of common source from horses in southwestern China (Sun et al., 2007). The former considers that the Chinese pony derived from Hipparion or a kind of short wild horse in ancient times, for example, the Yunnan Wild Horse. The latter insists that all horses came with ancient Qiang people when they migrated to the south. Neither of the theories has yet been tested by molecular genetic studies. In the present study, genetic diversity and population structure for all Chinese pony populations were analyzed using 13 microsatellites, and the phylogenetic relations between them were investigated as well, these conclusions were investigated to elucidate their origin and evolution and to develop strategies for conservation.

MATERIAL AND METHODS

Samples collection and DNA extraction

Blood samples were taken from five pony distribution areas in China, as shown in Figure 1. Forty Debao pony samples (DB), 25 Guizhou pony samples (GZ), 21 Ningqiang pony samples (NQ), 53 Sichuan pony samples (SC), and 35 Yunnan pony samples (YN), for a total of 174 samples representing five Chinese pony populations, were obtained. The samples were randomly collected, half males and half females, and all ponies were more than 5 years old and not closely related to each other. Withers height for all ponies was less than 106 cm. Each sample consisted of 8-10 mL whole blood and was collected from a jugular vein in

EDTA-coated tubes, and then transported to the laboratory in an ice box. Genomic DNA was extracted from all blood samples using standard protease K digestion and the TIANamp Genomic DNA kit (Tiangen Biotech Beijing Co., Ltd.). The DNA samples were stored at -20°C.

Figure 1. Areas of distribution of 5 Chinese pony breeds sampled in this study. Guizhou (GZ) - Ceheng county, Guizhou; Ningqiang (NQ) - Ningqiang county, Shanxi; Sichuan (SC) - Yuexi county, Sichuan; Yunnan (YN) - Pianbian county, Yunnan; Debao (DB) - Debao county, Guangxi.

Microsatellites

Thirteen microsatellite loci were used in the present study (TKY16, LEX064, HTG20, HTG 21, HTG28, UM002, UM012, UM016, UM018, UCDEQ440, UCDEQ465, ASB2, and ASB17) according to recommendations of the Food and Agriculture Organization (FAO) and the International Society for Animal Genetics (ISAG). The basic information of 13 microsatellite loci is shown in Supplementary Table 1. PCR amplification was performed on a TC-512 thermal cycler (Staffordshire, UK) using a 20-µL reaction volume with 50-100 ng template DNA, 10 µL 2X Taq PCR Master Mix (0.1 U/µL Taq polymerase, 500 µM each dNTP, 20 mM Tris-HCl, pH 8.3, 3 mM $MgCl_2$, and 1.0 µL each primer (10 pM). Thermal conditions were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 45 s, 45 s at annealing temperature, and 1 min at 72°C, with a final extension at 72°C for 10 min. The PCR products were electrophoresed using an 8% urea polyacrylamide denaturing gel, and a 20-bp DNA ladder marker (TaKaRa Biotechnology Co., Ltd.) was used to estimate the size of the amplified products. The fragment sizes were visualized by silver staining according to Bassam (1991).

Data analysis

For each of marker and a single pony population, the POPGENE1.32 software (Francis et al., 2000) was used to estimate the basic variation indices, including the allele frequencies, observed number of alleles (N_A), effective number of alleles (N_E), observed hetero-

zygosity (H_O) and effective heterozygosity (H_E). Genetic differentiation between and within populations was estimated based on unbiased F-statistics according to Weir and Cockerham (1984), where the allelic richness (*AR*) for each locus and population was computed using the FSTAT 2.9.3 program (Goudet, 2001). Allelic richness was calculated using a rare sample size of 21 diploid individuals per population. The exact probability test was performed to determine departure from Hardy-Weinberg equilibrium (HWE) using GENEPOP 3.4 (Raymond and Russet, 1995). The exact P value was estimated from the Markov chain algorithm based on 10,000 dememorization steps, 20 batches and 5000 iterations per batch.

Nei's genetic distance (D_A; Nei et al., 1983) between pairs of populations and the neighbor-joining (NJ) tree between breeds were generated with the POPULATIONS package (Langella, 2002). The phylogenetic tree was constructed based on D_A; the robustness of the dendrogram was ensured using a bootstrap test of 1000 resamplings of loci. The TREEVIEW program (Page, 1996) was used to visualize the tree.

Finally, the STRUCTURE 2.1 software (Pritchard et al., 2000) was used to reveal clusters of individuals on the basis of their genotypes at multiple loci using a Bayesian algorithm. We used an admixture model, in which an individual may have mixed ancestry. The putative number of sub-populations (*K*) was from 2 to 5. We made 20 independent runs for each K to evaluate stability, with 10,000 repetitions in the burning period and then 100,000 repetitions using the Markov chain Monte Carlo (MCMC) method.

RESULTS

Microsatellite loci and genetic diversity

All microsatellite loci used in this study were successfully amplified with the right size and all showed high allelic variation in the five populations (Table 1). A total of 144 alleles were observed at 13 loci in 174 samples. The most variability across the five breeds was at UM018, where N_A, N_E, H_O, H_E, and *AR* were 16, 13.62, 0.81, 0.93, and 14.40, respectively. The least variability occurred at TKY16, where N_A, N_E, H_O, H_E, and *AR* were 8, 5.97, 0.70, 0.83, and 7.74, respectively. The mean N_A, N_E, H_O, H_E, and *AR* over all loci were 11.08, 9.12, 0.77, 0.88, and 10.28.

Table 1. Basic genetic parameters and F-statistics for the 5 Chinese pony breeds at 13 microsatellite loci.

Locus	N_A	N_E	H_O	H_E	*AR*	F_{ST}	F_{IS}
TKY16	8	5.97	0.70	0.83	7.74	0.070***	0.110**
LEX016	11	10.29	0.73	0.90	10.73	0.73***	0.140***
HTG20	10	9.45	0.80	0.89	9.85	0.064***	0.115*
HTG21	9	6.13	0.71	0.84	8.04	0.032***	0.123***
HTG28	11	8.84	0.83	0.89	10.56	0.142***	-0.063
UM002	12	10.00	0.63	0.90	11.20	0.044***	0.271***
UM012	11	8.62	0.86	0.89	10.14	0.028***	0.000
UM016	13	10.36	0.75	0.91	12.00	0.075***	0.115***
UM018	16	13.62	0.81	0.93	14.40	0.053***	0.090**
UCDEQ440	11	7.85	0.84	0.88	8.99	0.034***	0.006
UCDEQ465	15	13.66	0.83	0.93	13.94	0.054***	0.064*
ASB2	8	6.90	0.78	0.86	7.90	0.047***	0.054
ASB17	9	6.83	0.72	0.86	8.10	0.058***	0.110**
Mean	11.08	9.12	0.77	0.88	10.28	0.060***	0.087**

N_A = number of alleles; N_E = number of effective alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; *AR* = allelic richness; F_{ST} = fixation index resulting from the comparison of subpopulations to total population; F_{IS} = fixation of subpopulation. Significant levels of heterozygotes: *P < 0.05; **P < 0.01; ***P < 0.001.

The polymorphisms within the pony populations are listed in Table 2. N_A and N_E ranged from 6.78 and 5.38 in GZ to 9.15 and 6.78 in SC. H_O ranged from 0.66 in NQ to 0.82 in YN, and H_E ranged from 0.82 in GZ to 0.85 in DB and SC. *AR* was the highest in SC (8.52) and lowest in GZ (6.73). There were 7 private alleles discovered over all populations: two in NQ, SC and YN, one in GZ, and 34 shared alleles were observed in the five pony populations (data not shown).

Table 2. Population genetic diversity in 5 Chinese pony breeds analyzed using 13 microsatellite loci.

Breeds	Sample size	N_A	N_E	H_O	H_E	*AR*	F_{IS}
DB	40	7.85	6.51	0.79	0.85	7.74	0.070***
GZ	25	6.78	5.38	0.78	0.82	6.73	0.044*
NQ	21	7.23	5.87	0.66	0.84	7.23	0.219***
SC	53	9.15	6.78	0.77	0.85	8.52	0.096***
YN	35	8.08	6.16	0.82	0.84	7.83	0.030
Mean	174	7.82	6.14	0.76	0.84	7.61	0.092**

N_A = number of alleles; N_E = number of effective alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; *AR* = allelic richness; F_{IS} = population inbreeding coefficient. Significant levels of heterozygotes: *P < 0.05; **P < 0.01; ***P < 0.001. DB = Debao; GZ = Guizhou; NQ = Ningqiang; SC = Sichuan, and YN = Yunnan.

The results of the exact P value test for HWE of all breed-locus combinations are shown in Supplementary Table 2; 26 (40%) of 65 loci for all five breeds showed significant deviation (P < 0.05) from HWE. Only the HTG28 and UCDEQ440 loci were in HWE in all 5 populations investigated, but all 5 populations deviated from HWE at some loci. On average, 5.2 loci per breed and 2 breeds per locus deviated significantly from HWE. The NQ pony had the most number of loci in disequilibrium (8 loci), followed by SC pony (6 loci), and 4 loci were in disequilibrium for the rest populations. In fact, only the YN population was in HWE, and the other populations deviated from HWE.

The inbreeding coefficients (F_{IS}) estimated for five pony breeds are shown in Table 3. The significant deviation of the F_{IS} values from zero, indicating heterozygosity deficiency was observed for nine loci in the NQ population; however, only two loci were observed in the DB population. Heterozygosity excess, reflected in the significant negative F_{IS} values was observed for TKY16 and UCDEQ440 in the GZ population, UCDEQ440 in the NQ population, and UM016, ASB2 and ASB17 in the YN population. The mean estimates of F_{IS} for the five pony breeds differed significantly from zero, except for the YN population.

Table 3. Within-population inbreeding estimates (F_{IS}) for Debao (DB), Guizhou (GZ), Ningqiang (NQ), Sichuan (SC), and Yunnan (YN) breed.

Locus	DB	GZ	NQ	SC	YN
TKY16	0.048	-0.214*	0.467***	0.159*	0.101
LEX064	0.062	-0.087	0.637***	0.129*	0.132
HTG20	0.069	0.060	0.223*	0.038	-0.040
HTG21	-0.105	0.300**	0.241*	-0.046	0.446***
HTG28	-0.126	0.026	-0.158	0.023	-0.144
UM002	0.275***	0.345***	0.084	0.302***	0.280***
UM012	0.079	-0.031	0.159*	-0.039	-0.114
UM016	0.387***	0.080	-0.087	0.227***	-0.230**
UM018	-0.079	0.053	0.781***	-0.060	0.139
UCDEQ440	0.042	-0.180*	-0.200*	0.075	0.108
UCDEQ465	0.004	0.367***	0.040	0.019	0.000
ASB2	0.106	-0.074	0.427***	0.096	-0.192**
ASB17	0.101	-0.116	0.236*	0.354***	-0.155*
Mean	0.070***	0.044*	0.219***	0.096***	0.030

Significant levels of heterozygotes: *P < 0.05; **P < 0.01; ***P < 0.001.

The mean F_{ST} value (0.060) for the five pony breeds indicated that about 6% of the total genetic variation resulted from genetic differentiation between breeds, with the remaining 94% corresponding to differences within each breed. The pairwise F_{ST} values between the five breeds are shown in Table 4, with the values ranging from 0.0424 for the SC-YN pair to 0.0833 for the GZ-SC pair. All pairwise F_{ST} comparisons showed a significant value, which means the existence of genetic differentiation between all the population pairs. Thus, from 4.24 to 8.32% of the microsatellite variability is explained by the subdivision of the population.

Table 4. Nei's D_A distance below the diagonal and pairwise population differentiation (F_{ST}) above the diagonal among five pony breeds.

Breeds	DB	GZ	NQ	SC	YN
DB	-	0.0621**	0.0549**	0.0551**	0.0542**
GZ	0.3624	-	0.0590**	0.0833**	0.0806**
NQ	0.3332	0.3280	-	0.0541**	0.0696**
SC	0.3384	0.4968	0.3494	-	0.0424**
YN	0.3323	0.4596	0.4505	0.2879	-

Significant levels of genetic differentiation: **P < 0.01. For abbreviations, see Table 2.

Breed relationship

The allele frequencies from 13 microsatellites were used to generate the D_A genetic distance for each pair of the five pony populations (Table 4). The D_A genetic distance ranged from 0.2879 between SC-YN ponies to 0.4968 between GZ-SC ponies. A neighbor-joining tree was constructed based on D_A genetic distance with relatively high bootstrap values. As shown in Figure 2, the five Chinese pony populations separated into three different clusters: DB showed the farthest genetic distance from the other four pony populations as the first cluster, the GZ and NQ comprised the second clustered with a bootstrap value of 56%, and SC and YN made up the third cluster with a bootstrap value of 97%.

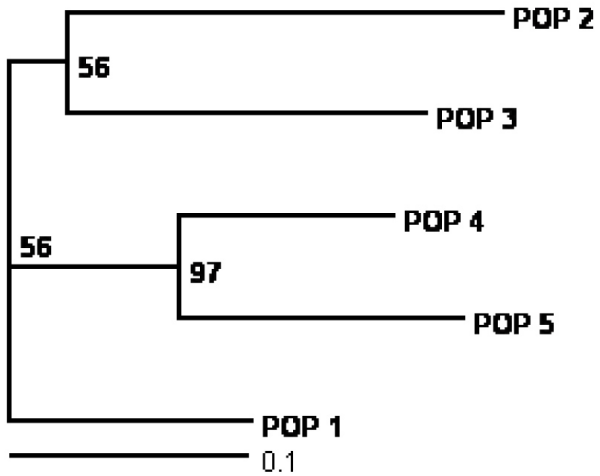

Figure 2. The neighbor-joining tree for 5 Chinese pony breeds based on Nei's unbiased D_A distances (Nei et al., 1983) analyzed by microsatellite markers. The numbers at nodes was calculated with 1000 bootstrap replications. Scale bar represents the branch length. POP 1 = Debao; POP 2 = Guizhou; POP 3 = Ningqiang; POP 4 = Sichuan; POP 5 = Yunnan.

In Figure 3, each individual is represented by a thin vertical line, which is divided into K colored segments representing the fraction of each individual belonging to each inferred cluster. A black line separates individuals of the different populations. Bayesian structure analysis did not provide an unequivocal number of clusters. However, when *K* = 3, most individuals belonging to the five pony breeds were assigned to their respective clusters (Supplementary Table 3). Thus, the optimum clustering value was K = 3.

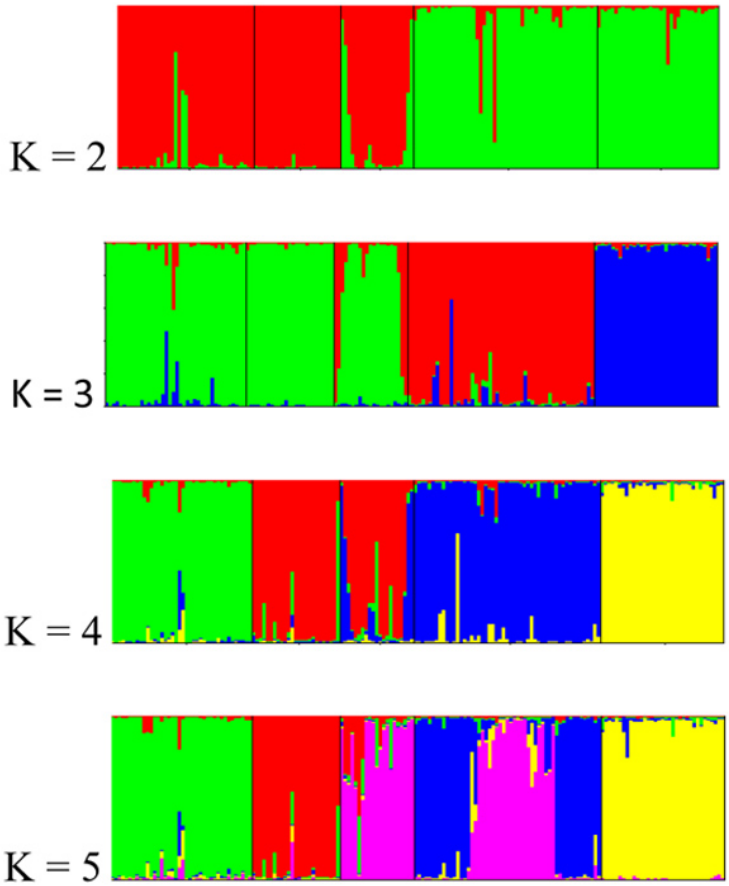

Figure 3. Structure analysis of five Chinese pony breeds assuming K = 2, 3, 4, and 5. Each pony is represent by a single vertical line that is divided into colored segment, representing the individual's membership in the cluster of the corresponding color and the lengths proportional to the estimated membership of the inferred cluster.

DISCUSSION

Microsatellites are widely adopted to quantify genetic variation within and between breeds and to assign individuals to reference populations (Bruford et al., 2003; Rosenberg et al., 2003). Most studies have been carried out on genetic diversity in horses (Glowatzki-Mullis et al., 2005; Behl et al., 2007; Felicetti et al., 2010; Ling et al., 2011). Some authors have uti-

lized microsatellites to analyze genetic diversity of Chinese ponies (Zhang et al., 2008, 2009), but the genetic diversity and population structure of all five Chinese pony populations in the present work have not been reported based on microsatellite markers.

Within the framework of breed conservation, genetic characterization is important in maintaining breed integrity and is a prerequisite for managing genetic resources (Bjornstad and Roed, 2002). The level of allelic richness, the number of alleles, the average number of alleles for all loci in the population and heterozygosity are used to measure genetic diversity. In the present study, the average number of alleles observed per locus over all populations was 11.08, which means that the microsatellites we selected are suitable for genetic diversity studies according to the suggestions by Barker (1994). Considering all populations, the same locus can have a different number of alleles, reflecting the difference in genetic diversity in populations. The value of heterozygosity is considered a better estimator of the genetic variability present in a population (Nei and Kumar, 2000). Our current study revealed a greater average genetic diversity level (H_E = 0.84) compared to previous reports of the GZ pony (H_E = 0.769) (Zhang et al., 2009) and the NQ pony (H_E = 0.78) (Zhang et al., 2008). The average H_E for GZ and NQ was 0.82 and 0.84, higher than that in the study by Zhang et al. (2008, 2009) (0.769, 0.78), this is probably because of the differences in sampling place, the methods in genotyping, and so on. *AR* is a measure of the number of alleles that can be corrected to be independent from the sample size (Petit et al., 1998). The average level of *AR* (7.16) was similar to previous findings reported in Chinese indigenous horse breeds (7.21) (Ling et al., 2011). The higher genetic diversity levels present in the five Chinese pony populations may be the result of low rate of selection pressure owing to the lack of improvement programs and the existing genetic lineages within pony populations.

Results of the exact P value test and mean F_{IS} values indicated that all investigated populations deviated from HWE except the YN breed. There are many reasons for disequilibrium such as selection, migration, mutation, and inbreeding. In addition, null alleles may be another reason for a population deviating from HWE. In fact, when allelic diversity is high and sample size is moderate, a significant deviation from HWE (Guo and Thompson, 1992), was caused in our study. The main reasons may be because all the samples were taken from villages; there were no selective program and breeding schemes carried out in these areas, the genetic base of ponies was wide and inevitable inbreeding occurred frequently. The closer the H_O is to H_E, the smaller effects of external selection and inbreeding are on this breed, which shows that the population may be in a HWE state. This was the case for the YN breed, where the H_O was close to H_E. However, the F_{IS} values of the YN breed had three loci with significant heterozygosity deficiency and heterozygosity excess in the other three loci, signifying that its gene pool could have undergone rapid unexpected changes, which may lead to inbreeding depression.

The F_{ST} values obtained were moderate, where a small mean F_{ST} value (0.060) was detected in the five pony populations, suggesting that most differences existed within populations. In general, our current data reveal a higher level in comparison with previous reports of Chinese horses (0.024) (Ling et al., 2011). However, the F_{ST} value was smaller than that of other studies, such as 0.078 for seven Spanish breeds (Canon et al., 2000), 0.100 for some Polish breeds (Zabek et al., 2005) and 0.099 for French horses (Leroy et al., 2009). The results of pairwise F_{ST} were equal to the genetic distance, indicating that the most genetic diversity existed in the GZ-SC pair and the least genetic diversity existed in the SC-YN pair.

The NJ method was used to construct a phylogeny tree in this study. However, the

tree showed that the branches did not correspond to the geographic location; in other words, we did not find geographic distance in proportion to population genetic similarity. Population determination is usually based upon geographic origin of samples or phenotypes; however, the genetic structure of population is not always reflected in the geographic proximity of individuals (Evanno et al., 2005). In this study, this might have been caused by gene flow between the pony populations and the fact that the pony populations in villages have low selection pressure.

Because likelihood maximization intrinsically favors partitions with more clusters, the *lnP*(*D*) did not provide an unequivocal number of clusters. However, when individual ponies were clustered assuming the number of breeds to be three, 94.4% of individuals belonging to the DB pony population were assigned to the first cluster, 91% of GZ and 78.1% of NQ assigned to the second cluster, and finally, 92.9% of SC and 91.8% of YN assigned to the third cluster. The results are inline with the N-J phylogeny tree.

In conclusion, the genetic analysis in the current study showed that all of five Chinese pony populations have a high level of genetic diversity. They are more likely to be able to cope with future challenges. Chinese pony populations are conserved in natural ways, and these findings should aid in developing efficient conservation/breeding strategies for Chinese pony breeds.

ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of China (NSFC) (#30960162). The authors acknowledge the assistance of five animal institutes in China with sample collection.

REFERENCES

Barker JSF (1994). A global protocol for determining genetic distances among domestic livestock breeds. Proccedings of the 5th World Congress of Genetic and Applied Livestock Production, Guelph, 21: 501-508.

Bassam BJ, Caetano-Anolles G and Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80-83.

Behl R, Behl J, Gupta N and Gupta SC (2007). Genetic relationships of five Indian horse breeds using microsatellite markers. *Animal* 1: 483-488.

Bjornstad G and Roed KH (2002). Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. *Anim. Genet.* 33: 264-270.

Bruford MW, Bradley DG and Luikart G (2003). DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4: 900-910.

Canon J, Checa ML, Carleos C, Vega-Pla JL, et al. (2000). The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Anim. Genet.* 31: 39-48.

Evanno G, Reganut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-1620.

Felicetti M, Lopes MS, Verini-Supplizi A, Machado AC, et al. (2010). Genetic diversity in the Maremmano horse and its relationship with other European horse breeds. *Anim. Genet.* 41 (Suppl 2): 53-55.

Francis CY, Yang RC and Tim B (2000). POPGENE (1.32). Available at [http://www.ualberta.ca/~fyeh/popgene_download.html]. Accessed October 18, 2011.

Glowatzki-Mullis ML, Muntwyler J, Pfister W, Marti E, et al. (2005). Genetic diversity among horse population with a special focus on the Franches-Montagnes breed. *Anim. Genet.* 37: 33-39.

Goudet J (2001). FSTAT. A Program to Estimate and Test Gene Diversity and Fixation Indices (Version 2.9.3.2). Available at [http://www2.unil.ch/popgen/softwares/fstat.htm]. Accessed November 23, 2011.

Guo SW and Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.

He XR (1987). Pony. *Bull. Vet. Coll. PLA* 7: 213-216.

Jiang Q, Wei Y, Huang Y, Jiang H, et al. (2011). The complete mitochondrial genome and phylogenetic analysis of the Debao pony (*Equus caballus*). *Mol. Biol. Rep.* 38: 593-599.

Langella O (2002). POPULATIONS (Version 1.2.31). Available at [http://www.bioinformatics.org/~tryphon/populations/]. Accessed January 6, 2012.

Leroy G, Callede L, Verrier E, Meriaux JC, et al. (2009). Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genet. Sel. Evol.* 41: 5.

Ling YH, Ma YH, Guan WJ, Cheng YJ, et al. (2011). Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Anim. Genet.* 42: 56-65.

Nei M and Kumar S (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nei M, Tajima F and Tateno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* 19: 153-170.

Page RD (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357-358.

Petit RJ, Ei-Mousadik A and Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.* 12: 844-855.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Raymond M and Rousset F (1995). GENEPOP (Version 1.2). population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.

Rosenberg NA, Li LM, Ward R and Pritchard JK (2003). Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* 73: 1402-1422.

Sun YJ, Ta N, Sai N and Mang L (2007). The utilize and conservation of genetic resource of China pony. *Heilongjiang Anim. Sci. Vet. Med.* 7: 14-16.

Wang ZS and Yue GF (2002). Preliminary study on blood genetic markers in chinese miniature horses. *Heilongjiang J. Anim. Sci. Vet. Med.* 10: 3-5.

Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Zabek T, Nogaj A, Radko A, Nogaj J, et al. (2005). Genetic variation of Polish endangered Bilgoraj horses and two common horse breeds in microsatellite loci. *J. Appl. Genet.* 46: 299-305.

Zhang T, Lu HC, Liang L and Ceng Z (2008). Genetic diversity of Ningqiang pony using microsatellite markers. *Hubei Agric. Sci.* 8: 868-870.

Zhang T, Lu HC and Li XS (2009). The genetic diversity of Guizhou pony using 4 microsatellite markers. *Acta Ecol. Anim. Domastici* 30: 19-21.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Characteristics of 13 microsatellite loci used in this study.

Locus	Chromosome and location	Primer sequences (forward and reverse)	Size ranges (bp)	Annealing temperature (°C)
TKY16	18	F: GGTTATGGTTTGGTATCTGTC R: AAAACAATGGCTTCCTGGTCA	124~153	57.6
LEX064	20	F: ACCCTTTCCGCAGACAA R: CACATCAGAGCCCATCTTCTC	203~253	59.8
HTG20	7	F: CTGGTTTACCTTCCTACAG R: CCAATGGTTCCTCTGAGAAG	138~186	58.0
HTG21	22	F: ATTACTTCCTCCAGGTATCTCAG R: AGGCAGGGCTGGGAGACGT	125~161	57.8
HTG28	18	F: AATCAACTAATATTAGGCCTCCT R: GAATACAGTTCTAGGGGCGT	164~211	56.2
UM002	1	F: AGTGGCAGCTAAAGATG R: TTTTGGTCCTTGTAGGAG	225~281	52.0
UM012	24	F: GGAATTTAGACTATGACTGAGG R: GCCACCTGAACACTTTAC	97~149	58.3
UM016	14	F: TTCCTCCACTATCTCTCCCTC R: GCAAAAATGCACAGCCTC	171~215	57.5
UM018	23	F: AGATAGGGAAAGAACGAAG R: GGTTAGTGTTCGGATTAGG	200~265	55.5
UCDEQ440	1	F: TGTTCGGACAGTGTGGAT R: GCAGGGTATGTGTGTGCT	97~137	56.0
UCDEQ465	6	F: AACCAGTCCCTACATAGAAC R: CTCACAACCAAGCATACA	188~265	54.4
ASB2	15	F: CCTTCCGTAGTTTAAGCTTCTG R: CACAACTGAGTTCTCTGATAGG	160~198	54.0
ASB17	2	F: GAGGGCGGTACCTTTGTACC R: ACCAGTCAGGATCTCCACCG	95~127	57.0

Supplementary Table 2. P value per population and loci in the Hardy-Weinberg test.

Locus	P value				
	DB	GZ	NQ	SC	YN
TKY16	0.1198	0.9879	0.0011	0.0020	0.2586
LEX064	0.0691	0.3446	0.0000	0.0063	0.0005
HTG20	0.0450	0.0135	0.0048	0.2271	0.3469
HTG21	0.9053	0.0116	0.0259	0.8195	0.0000
HTG28	0.3196	0.0513	0.8579	0.2163	0.9700
UM002	0.0000	0.0000	0.2072	0.0005	0.0001
UM012	0.0702	0.1545	0.0333	0.4581	0.9460
UM016	0.0000	0.1684	0.7768	0.0002	0.6744
UM018	0.8526	0.0726	0.0000	0.7728	0.0263
UCDEQ440	0.2715	0.8790	1.0000	0.1289	0.0883
UCDEQ465	0.1961	0.0007	0.3319	0.0967	0.3494
ASB2	0.0352	0.8534	0.0000	0.0008	1.0000
ASB17	0.0642	0.3523	0.0297	0.0000	0.8547

For abbreviations, see Table 2.

Supplementary Table 3. Proportion of membership of five pony populations in each of the 3 clusters.

Breed	Inferred clusters		
	1	2	3
DB	0.028	0.944	0.028
GZ	0.910	0.075	0.015
NQ	0.781	0.067	0.152
SC	0.051	0.020	0.929
YN	0.018	0.064	0.918

For abbreviations, see Table 2.

