Conservation Genetics (2006) 7:117–127
DOI 10.1007/s10592-005-8670-z

Phylogeography and origin of sheep breeds in Northern China

Yue-Hui Ma[1,*], Shao-Qi Rao[2,3,*], Shen-Jin Lu[1,4], Guan-Yu Hou[5], Wei-Jun Guan[1], Hong-Bin Li[1], Xia Li[3], Qian-Jun Zhao[1] & Jun Guo[1]

[1]*Division of Animal Resources and Conservation, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, 100094, China;* [2]*Departments of Molecular Cardiology and Cardiovascular Medicine, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio, 44195, USA;* [3]*Department of Bioinformatics, Harbin Medical University, Harbin, Heilongjiang, 150086, China;* [4]*Linyi Normal University, Linyi, Shandong, 276000, China;* [5]*College of Animal Science and Veterinary, Inner Mongolia Agricultural University, Huhhot, Inner Mongolia, 010018, China*

*(*Corresponding Authors: Phone: +86-10-62813463; +1-216-4440056; Fax: +1-216-4442682; E-mail: mayuehui@iascaas.net.cn; raos@ccf.org)*

Received 25 April 2005; accepted 8 June 2005

Key words: genetic distance, genetic diversity, microsatellite, phylogeography, sheep

Abstract

With the establishment of modern sheep production systems in China, various forms of hybridization with Western breeds and between native breeds have been utilized for genetic improvement. At the same time, the progressive destruction or deterioration of sheep habitat has accompanied urbanization in China. Together these factors have accelerated the loss of genetic diversity, or even resulted in the extinction of some indigenous breeds. It is therefore important that efficient strategies for surveillance, evaluation, conservation and utilization of available genetic resources are developed for this species. In this study, a total of 30 microsatellite markers were used to assess genetic diversity for 12 native breeds and one Western sheep breed in Northern China. The high polymorphism information contents at the 30 markers, varying from averages of 0.519 to 0.666 for the 13 breeds, imply the retention of natural variation from source populations in the domestic breeds from different geographic regions in China. Analysis of genetic differentiation revealed substantial divergence among these breeds. Neutrality tests indicated that more than one third of the 30 loci were in departure from neutrality, implying that some evolutionary forces (e.g. selection and migration) had acted on these populations. Phylogenetic and phylogeographic analyses displayed a remarkable degree of consistency between geographic origins, breeding histories and the pattern of genetic differentiation.

Introduction

China has a long history of sheep domestication. A total of 30 breeds consisting of 15 indigenous, 7 improved, and 8 Western breeds have been characterized based on their breeding histories, physiological characteristics, behaviors and geographic distributions (Zheng 1988). However, not all have been characterized, and there may be more than 40 in total (Zheng 1988). Most of the indigenous breeds are highly adapted to local environmental conditions. For example, some local breeds have the ability to deposit a large amount of fat in the body to meet nutritional demands during the winter and spring. The varieties found among indigenous breeds provide for some important 'niche' markets. For instance, Ningxia Tan is one of the major breeds for fur production and carpet wool (a type of coarse wool) in China and Small-(Fat)-Tail Han is one of the most prolific breeds on the world.

In order to establish modern sheep production systems, competitive for high market returns, a large number of breeds were imported into China

during the second half of the last century to initiate a nation-wide campaign for the genetic improvement of native breeds. However, there has been concern that this may have led to the loss of unique indigenous genetic information or the extinction of indigenous breeds through artificial selection for traits of economic importance. This is especially because of the known unfavorable genetic correlations between 'economic' and 'adaptability' traits (Rao and Notter 2000). About 15% of sheep and goat breeds have been classified as susceptible to loss of genetic diversity or extinction based on criteria proposed by (Simon 1999). The continual loss of genetic resources of domestic animals including poultry (Fulton and Delany 2003) is a worldwide trend, especially in developing countries, due to the development and establishment of more efficient livestock production systems (i.e., raising a few high productive breeds or hybrids) for the competitive domestic and world markets.

Conservation practice is based on the need to maintain genetic variation, which retains deleterious recessive mutations in a heterozygous state and provides adaptive potential in a changing environment and market demands. During the last century, genetic diversity loss in sheep occurred in an unprecedented pace in China, with the wide use of crossbreeding and with the progressive destruction or deterioration of the ecological environments for sheep habitat in China. There is thus a great demand to quantifying the amount of genetic variation in the current populations to permit genetic information to be used in sheep conservation practice. Of the many types of genetic markers now available, private microsatellite loci are best suited for answering several conservation issues (Guti'errez-Espeleta et al. 2000). First, a genetic survey can directly measure the amount of genetic variation in sheep. Second, genetic data at the molecular level can be instrumental for clarifying the ambiguities in the current taxonomic classification based on morphological and physiological data, which is deemed to be insufficient for determining relationships between breeds (Arranz et al. 1998). Third, recognizing historical patterns of genetic variation among the sheep breeds is required to preserve evolutionary relationships for conservation. For example, gene flow between different lines within the same breed has been a valuable part of sheep conservation efforts (Crandall et al. 2000). But, combining genetically different populations of sheep could alter adaptations to local environments and subsequently lower the fitness of populations (Guti'errez-Espeleta et al. 2000).

The issues for the conservation of domestic animals and genetic resources have been well recognized (Chen 1995; Chen and Cao 2001). Hence, we have established a large-scale gene bank consisting of cryopreserved animal embryos and semen, DNA and tissues for the endangered domestic animal breeds in China (Feng et al. 1997; Li et al. 2000; Fan et al. 2002; Ma et al. 2003b). In this study, we used a total of 30 polymorphic microsatellite markers to assess the genetic diversity among 12 native breeds and one Western sheep breed in Northern China.

Materials and methods

Subject recruitment and geographic distribution

Strategies for collecting DNA samples were based on the approach proposed in (Chang et al. 1989), i.e., randomly sampling from a typical flock in the central production region for each breed. The 13 breeds in the present study (Table 1) consist of well investigated and documented breeds (e.g. Ujumiqin, WU; Small-(Fat)-Tail Han, XH; and Tong, TY), indigenous breeds with unique characteristics that are less well investigated (e.g. Qiaoke, QK; Ganjia, GJ; and Oula, OL), one endangered breed (Langzhou Large-(Fat)-Tail Han, LD) and one breed with high risk for extinction (Hanzhong, HZ). The Western breed (Poll Dorset, PD, imported from Australia) was used for comparison. The studied breeds cover most of the available breeds in five provincial administrative regions (Inner Mongolia, Ningxia, Shaanxi, Gansu and Beijing). The origins and breeding histories of the studied breeds are diverse, including four breeds of the Tibetan origin (OL, GJ, QK; and Ming Black, MQ), five breeds of the Mongolia origin (SU, WU, XH, TY; and Tan, TA), one breed of the Middle Eastern origin (LD) and one breed of the England origin (PD). We also included an improved Mongolia breed (Inner Mongolian Wool, NE) that had genetic contributions from the local breeds, Soviet Merino and Australian Merino in its breeding history. Yet, the

Table 1. The geographic locations and the numbers of samples recruited for each sheep breed

	Breed	Location	No. of samples
1	Sonid (SU)	Sonid, Inner Mongolia, China	44
2	Ujumqin (WU)	East Ujumqin, Inner Mongolia, China	44
3	Inner Mongolia Wool (NE)	Zhenglianqi, Inner Mongolia, China	43
4	Tan (TA)	Yanchi, Ningxia, China	44
5	Small-Tail Han (XH)	Shunyi, Beijing, China	39
6	Hanzhong (HZ)	Hanzhong, Shaanxi, China	44
7	Tong (TY)	White Water, Shaanxi, China	44
8	Langzhou Large-Tail Han (LD)	Lanzhou, Gansu, China	44
9	Ming Black (MQ)	Ming, Gansu, China	44
10	Oula (OL)	Maqu, Gansu, China	44
11	Ganjia (GJ)	Xiahe, Gansu, China	44
12	Qiaoke (QK)	Maqu, Gansu, China	44
13	Poll Dorset (PD)	Beijing, China	23
Total			545

origin(s) of a studied breed (HZ) remains controversial, based on the previous studies of its morphological data, physiological characteristics and historical evidence (Sun et al. 2003b).

Ear-tip tissue samples from 545 sheep were collected from 13 different locations (Figure 1) across the five regions. It should be noted that two breeds were not sampled from their central production regions. PD sheep were from the research farm affiliated with the Institute of Animal Science, Chinese Academy of Agricultural Sciences, and XH were from a crop region in Beijing. Between 39 and 44 samples were collected for each breed with the exception of PD, for which only 23 samples were obtained (Table 1). The protocol for bio-sample recruitment was made in accordance with the relevant laws to conservation and animal welfare in China and was approved by each participating center's institutional review board.

Genotyping

About 0.5 g ear-tip tissue was obtained by a sterilized ear punch and then put into a centrifuge tube with 70% ethanol for cryopreservation under −20 °C. Genomic DNA was extracted using the method of phenol extraction (Sambrook et al. 1989). Thirty microsatellite markers (Table 2) were chosen from two public databases (http://www.marc.usda.gov and http://www.ncbi.com) and some were originally provided by the Italian Institute of Animal Genetic Resource Conservation and Utilization. The 30 microsatellite markers are distributed on 19 chromosomes of the sheep genome (Crawford et al. 1994, 1995), without close linkage between any two loci. There were at least 4 allelic variants in each breed at each of the 30 loci (Barker 1994). Primers for each bovine marker (de Gortari et al. 1997) and others (Table 2) were designed and made by SBSBIO Inc. (Beijing, China). The multiplex PCR amplifications were carried out in a total volume of 12 μl containing 0.2 mmol L^{-1} of each deoxynucleotide triphosphate (dNTPs), 6 pmol of each primer, 1 unit of *Taq*-polymerase, 1.5 mmol L^{-1} $MgCl_2$ and 50 ng of sheep genomic DNA. Electrophoresis was performed for each product on 7% denaturing polyacrylamide gel for about 2–4 h, and then treated with 0.1% silver nitrate for half an hour to develop the bands. Alleles and their sizes were determined by comparison with the corresponding Cosmid clone of PCR products, as described previously (Barker 1994; Ma et al. 2003a).

Statistical analysis

Allele frequencies were estimated from the genotyping data of the co-dominant microsatellite markers by direct counting. We analyzed each locus in each breed, each locus across breeds and each breed across loci. Marker characteristics for each breed in terms of polymorphism information content (PIC) (Botstein et al. 1980), genetic diversity index (mean heterozygosity across the marker loci, D_E), the averaged number of alleles and the averaged number of effective alleles (N_E)

Figure 1. The geographic locations of the central inhabitations for the 13 sheep breeds (11 Chinese indigenous breeds, 1 improved breed and 1 Western breed). The abbreviations for the 13 breeds are: SU: Sonid; WU: Ujumqin; NE: Inner (Nei) Mongolia wool sheep; TA: Tan; XH: Small-(Fat)-Tail Han; HZ: Hanzhong; TY: Tong; LD: Langzhou Large-(Fat)-Tail Han; MQ: Ming Black; OL: Oula; GJ: Ganjia; QK: Qiaoke; and PD: Poll Dorset.

(Hartl and Clark 1989) were computed. D_E was used to measure the amount of genetic variation present in each breed using data from multiple genetic loci. To characterize and determine whether the selected marker loci are appropriate for phylogenetic inference, first, the locus-by-locus heterozygosity as a measure for the overall genetic diversity in all the breeds (H_T) or for a specific breed (H_S, within-population heterozygosity) was computed, using the proportion of heterozygotes in the referenced population; second, the coefficient of genetic differentiation (G_{ST}), the ratio of the inter-population genetic diversity over the genetic diversity for all populations (Nei 1972), was used to reveal genetic divergence between the breeds; third, empirical neutrality tests (Fu and Li 1993) were conducted to assess the influences of selection or other evolutionary forces for each loci using 1000 permutations. Next, a genetic tree analysis of the genotype data of all 30 loci using Neighbor-Joining (NJ) method with a bootstrap support of 1000 permutations (Sneath and Sokal 1973) and the Nei's distance (Nei 1972) was conducted to explore the genetic relationships between the breeds. Bootstrapping over the loci (1000 replicates), as implemented in the DISPAN software package (Ota 1993), tested the reliability of the nodes in the phylogenetic tree. Finally, a Mantel test was used for assessing the relationships between the partitioning of genetic diversity and geographic origin. This test involves measuring the association between the elements in two matrices by a suitable statistic, and then assessing the significance of this statistic by comparison with the distribution found by randomly reallocating the order of the elements in one of the matrices

Table 2. The PCR anneal temperatures, the chromosome assignments and the number of alleles for the 20 microsatellite markers used in the study

Marker	Chromosome	No. of alleles	Anneal temp. (°C)
BM4311	6	13	58
BM6444	2	7	52
URB037	2	14	60
MB066	11	9	59
BM315	–	9	56
MAF70	4	18	60
BL6	23	10	52
BMS1004	15	13	55
AGLA269	23	18	52
BMS1248	3	14	60
BM6404	–	12	58
BMS574	1	17	60
BMS1714	25	10	54
BMS1724	8	22	58
BMS875	19	13	54
MB067	22	13	58
MB009	9	12	53
BM3413	18	14	60
BM3501	3	14	58
BMS1341	2	9	59
BMS1678	9	14	55
BMS710	3	6	60
BM1227	8	9	51
BM1225	16	20	56
ILSTS021	–	5	54
BM203	26	22	58
MB023	25	14	52
BM3033	7	13	58
BMC1206	21	14	54
BM6526	26	12	57

(Bonnet and Van de Peer 2002). All the statistical analyses were performed with one or combinations of several genetic software: GDA (gene data analysis) (Weir 1996), DISPAN (Ota 1993), ARLEQUIN (Schneider et al. 2000), POPGENE (Yeh and Boyle 1997), and ZT, a software tool for simple and partial Mantel tests (Bonnet and Van de Peer 2002).

Results

Gene frequencies

A total of 390 alleles at the 30 microsatellite markers were detected in the 13 breeds, with an average of 13 alleles per locus. The number of alleles per locus ranged from 5 to 22 (Table 2). Marker BL6 could only be amplified in six breeds, and MB067 only in nine breeds, which may reflect the locus specificity for these loci in some breeds. Some alleles were shared by all 13 breeds, while others were specific to one or several breeds. MQ and XH sheep have private alleles which may be useful as biomarkers for breed origin (Li et al. 2002; Long et al. 2003). Most loci were polymorphic for all breeds, but some were monomorphic in specific breeds.

Genetic variabilities within breeds

The averaged PIC over the 30 markers for each breed ranged from 0.519 to 0.666. The averaged heterozygosity (D_E) over the 30 markers varied from 0.589 to 0.714 (Table 3). All loci satisfied the criterion of being highly polymorphic (PIC > 0.5) in each breed (Botstein et al. 1980). The averaged effective number of alleles at the 30 markers had a range of 2.2–3.7 in the 13 breeds and all breeds had an average number of alleles greater than four (Table 3).

Analysis of microsatellite diversity, differentiation and neutrality

The total genetic heterozygosity (H_T), average heterozygosity within a population (H_S) and coefficient of gene differentiation (G_{ST}) at each of the 30 microsatellite markers were estimated using the DISPAN package (Table 4). H_T ranged from 0.466 to 0.823, with an average of 0.721 and H_S ranged from 0.325 to 0.740, with an average of 0.624, indicating that the selected molecular markers provided adequate information for sheep phylogenetic inference. The coefficient of gene differentiation for the 30 loci varied in a small range from 0.071 to 0.229, with an average of 0.132, implying that these breeds had undergone substantial genetic divergence.

Neutrality tests for each locus were performed with 1000 simulations, as implemented in POPGENE. An empirical significance ($P < 0.05$) of departure from the neutrality hypothesis was claimed if the observed F value was outside the simulated 95% confidence interval. Over half of the loci (a total of 19 markers: BM4311, URB037, MB066, BM315, MAF70, BL6, AGLA269,

Table 3. Measures for genetic variability at 30 microsatellite loci for each breed

Breed	PIC	D_E	No. of alleles	N_E
SU	0.610	0.665	4.3	3.0
WU	0.649	0.704	4.4	3.3
NE	0.587	0.677	4.5	3.1
TA	0.569	0.695	5.0	3.7
XH	0.643	0.697	4.0	3.2
HZ	0.666	0.714	4.9	3.6
TY	0.614	0.686	4.3	3.2
LD	0.590	0.658	5.1	3.2
MQ	0.634	0.696	5.4	3.4
OL	0.601	0.680	5.2	3.1
GJ	0.617	0.593	4.6	2.2
QK	0.644	0.710	5.4	3.5
PD	0.519	0.589	4.3	3.0

PIC, polymorphism information content.D_E, genetic diversity index, mean heterozygosity over the 30 markers for each breed.N_E, the averaged effective number of alleles over the 30 markers for each breed.

BM6404, BMS875, MB067, MB009, BM3413, BM3501, BM1341, BMS710, BM1227, ILSTS021, MB023 and BM3033) were consistent with the neutrality hypothesis (Table 4). However, we did not find a significant difference in phylogenetic and phylogeographic analyses when non-neutral loci were omitted (data not shown).

Genetic relationships between the breeds

The Nei's distance (D_A) (Nei 1972; Nei et al. 1983) between the 13 sheep breeds were computed using the DISPAN package and were derived using all the 30 microsatellite loci considered together (values are shown in Table 5). The genetic distance between GJ and MQ both in Gansu province ($D_A = 0.098$), was the smallest; next was the one between SU and WU both in the Inner Mongolia ($D_A = 0.112$). PD, a Western breed, had the largest genetic distance in comparison with Chinese native breeds, as we expected. NJ trees (Figure 2) were generated using the D_A distance measure and were evaluated with bootstrap support based on 1000 permutations, as implemented in GDA (Weir 1996). The phylogenies indicate that several breeds (GJ, MQ, QK, LD, OL, TY and HZ) from the neighboring provinces of Gansu and Shaanxi (Figure 1) were well supported in a single lineage, while the breeds from Inner Mongolia (SU, WU and NE) were less well supported in a separated lineage.

Partitioning of genetic diversity and geographic origin

The NJ phylogenetic analysis demonstrates a relationship between the partitioning of genetic diversity and geographic origin, as the two broadly defined lineages clearly represent different geographic clusters (Figure 1), with the exception of TA. We used Mantel tests to assess the relationships between genetic (pairwise $F_{ST}/(1-F_{ST})$) and geographic (ln(km)) distance matrices using ZT, a software tool for Mantel tests (Bonnet and Van de Peer 2002). The matrix for pairwise F_{ST} values between the breeds, calculated using ARLEQUIN, is given in Table 5 and its pattern among the breeds resembles remarkably the Nei's D_A matrix. Calculation of the geographic distance matrix involved calculation of centroids (average coordinates among data points) for each sampling location, followed by the calculation of distances between centroids. As the motherland of PD sheep was England, we omitted this breed from the Mantel test comparisons. The phylogeographic analyses of these breeds based on the genotype data displayed significant phylogeographic structure (the Mantel $r = 0.496$, $P = 0.0026$). Overall, these data establish a remarkable degree of consistency between geographic origins, breeding histories and the pattern of genetic differentiation.

Table 4. Locus-by-locus genetic diversity, genetic differentiation and neutrality

Marker	H_S	H_T	G_{ST}	Neutrality test	
				Observed *F*	Simulated *F* (95% CI)
BM4311	0.681	0.736	0.071	0.160	0.316 (0.159–0.640)
BM6444	0.542	0.648	0.168	0.193	0.502 (0.245–0.893)
URB037	0.652	0.729	0.106	0.140	0.269 (0.139–0.548)
MB066	0.622	0.714	0.129	0.199	0.403 (0.198–0.808)
BM315	0.470	0.596	0.216	0.218	0.427 (0.212–0.826)
MAF70	0.692	0.793	0.130	0.131	0.260 (0.131–0.502)
BL6	0.325	0.352	0.036	0.540	0.364 (0.180–0.719)
BMS1004	0.616	0.788	0.215	0.112	0.322 (0.155–0.677)
AGLA269	0.718	0.794	0.095	0.123	0.234 (0.123–0.480)
BMS1248	0.687	0.823	0.166	0.078	0.204 (0.110–0.415)
BM6404	0.618	0.560	0.190	0.297	0.365 (0.172–0.737)
BMS574	0.647	0.811	0.198	0.087	0.246 (0.122–0.506)
BMS1714	0.626	0.560	0.155	0.146	0.401 (0.196–0.787)
BMS1724	0.754	0.747	0.109	0.078	0.193 (0.104–0.409)
BMS875	0.697	0.847	0.103	0.155	0.305 (0.154–0.611)
MB067	0.530	0.778	0.116	0.219	0.303 (0.157–0.611)
MB009	0.661	0.598	0.111	0.177	0.345 (0.174–0.720)
BM3413	0.664	0.741	0.152	0.158	0.303 (0.155–0.573)
BM3501	0.681	0.782	0.070	0.185	0.309 (0.151–0.620)
BM1341	0.571	0.735	0.099	0.213	0.434 (0.195–0.827)
BMS1678	0.688	0.631	0.115	0.125	0.326 (0.164–0.655)
BMS710	0.520	0.781	0.153	0.249	0.496 (0.243–0.906)
BM1227	0.484	0.621	0.153	0.335	0.425 (0.205–0.828)
BM1225	0.661	0.568	0.131	0.092	0.209 (0.114–0.417)
ILST021	0.361	0.782	0.139	0.463	0.619 (0.302–0.962)
BM203	0.735	0.466	0.228	0.081	0.213 (0.112–0.406)
MB023	0.613	0.817	0.100	0.170	0.341 (0.166–0.700)
BM3033	0.588	0.696	0.120	0.177	0.342 (0.175–0.690)
BMC1206	0.635	0.686	0.156	0.098	0.286 (0.145–0.607)
BM6526	0.740	0.732	0.071	0.114	0.350 (0.170–0.712)
Mean	0.624	0.721	0.132		

Heterozygosity as a measure for genetic diversity for a specific breed (H_S, within-population heterozygosity) or for the overall genetic diversity in all the breeds (H_T) is the proportion of heterozygotes over the total number of genotypes in the referenced population. Coefficient of genetic differentiation (G_{ST}) was computed to be the ratio of the inter-population genetic diversity over the genetic diversity in the total populations. An empirical significance ($P < 0.05$) of departure from the neutrality hypothesis was claimed if the observed *F* value was outside the simulated 95% confidence interval (95% CI).

Discussion

Phylogenetic and phylogeographic analyses of the breeds have generally agreed with each other to support the conclusion that the partitioning of genetic diversities of the breeds is consistent with their geographic distributions. The data that four breeds of Tibetan origin (OL, GJ, QK and MQ) were clustered with LD on the NJ phylogenetic tree may be attributed to the fact that all the five breeds lived in a small region in the Gansu province (also see Figure 1). It is possible that gene flow may occur more frequently between these regional populations. Although HZ and TY in Shaanxi province were classified into the Mongolia group in the previous literature (Zheng 1988), they were clustered into a group with three Tibetan breeds instead. This phylogenetic result might also be explained by gene flow between the regional breeds.

Table 5. Genetic distances (D_A, below the diagonal) and pairwise F_{ST} (above the diagonal) between the breeds

	SU	WU	NE	TA	XH	PD	LD	OL	GJ	QK	MQ	HZ	TY
SU		0.101	0.316	0.264	0.228	0.389	0.329	0.271	0.289	0.259	0.265	0.293	0.284
WU	0.112		0.252	0.226	0.189	0.301	0.293	0.267	0.255	0.213	0.242	0.260	0.252
NE	0.326	0.215		0.301	0.276	0.400	0.323	0.314	0.298	0.263	0.264	0.301	0.275
TA	0.239	0.192	0.255		0.175	0.342	0.293	0.305	0.294	0.262	0.289	0.275	0.286
XH	0.207	0.159	0.234	0.144		0.328	0.302	0.287	0.293	0.237	0.278	0.269	0.262
PD	0.377	0.261	0.354	0.299	0.282		0.368	0.351	0.357	0.325	0.347	0.315	0.334
LD	0.303	0.258	0.285	0.259	0.262	0.325		0.178	0.144	0.154	0.171	0.185	0.186
OL	0.236	0.227	0.274	0.267	0.243	0.305	0.148		0.193	0.168	0.158	0.159	0.165
GJ	0.257	0.217	0.259	0.256	0.250	0.311	0.118	0.161		0.152	0.121	0.192	0.182
QK	0.228	0.179	0.227	0.227	0.199	0.283	0.128	0.139	0.125		0.145	0.167	0.153
MQ	0.231	0.209	0.232	0.254	0.240	0.310	0.145	0.131	0.098	0.120		0.157	0.146
HZ	0.256	0.225	0.263	0.240	0.231	0.275	0.156	0.132	0.162	0.140	0.132		0.139
TY	0.245	0.216	0.240	0.251	0.224	0.293	0.159	0.137	0.153	0.127	0.123	0.116	

In the other main lineage of the NJ tree, we observed that the only Western breed (PD) clustered with SU, a Mongolian breed, which may be explained by a recent history of genetic improvement for this breed. A large-scale crossbreeding campaign for genetic improvement of the native breed in the Inner Mongolia was undertaken in the last 20 years, although this practice had to be stopped because of the poor adaptability of the resulted crossbreds (mostly with Western wool breeds) to the harsh arid Mongolian plateau. The Inner Mongolia wool breed, NE, was formed in the last decade by crossing with Australian Merino rams, and was identified in the same broad lineage as PD and SU. This phylogenetic outcome is consistent with historical evidence that some of these breeds may share Merino genes. Dorset (the ancestor of PD) was formed in England centuries ago, by crossbreeding between the Merino sheep and the Horned Sheep of Wales (http://www.ansi.okstate.edu/breeds/sheep/). The occurrence of WU and SU in the same lineage is consistent with previous work based on morphological and physiological similarities, and with their breeding and evolutionary history (Zheng 1988). An alternative and plausible explanation is that their genetic similarity reflects recent gene flow due to their geographic proximity. The data showing that XH (a breed from agricultural regions in the central plains of China) clustered with breeds from the Mongolian plateau are also consistent with the breeding history of the breed. XH sheep originated from Mongolian sheep and were translocated to their current geographic range. However, TA is one of the few clear exceptions to the correlation between genetic similarity and geographic proximity. Although it is geographically closer to four breeds of Tibetan origin (GJ, OL, QK and MQ), TA was located on the lower lineage of the NJ tree (see also Figure 2), consisting of Mongolia breeds and the only

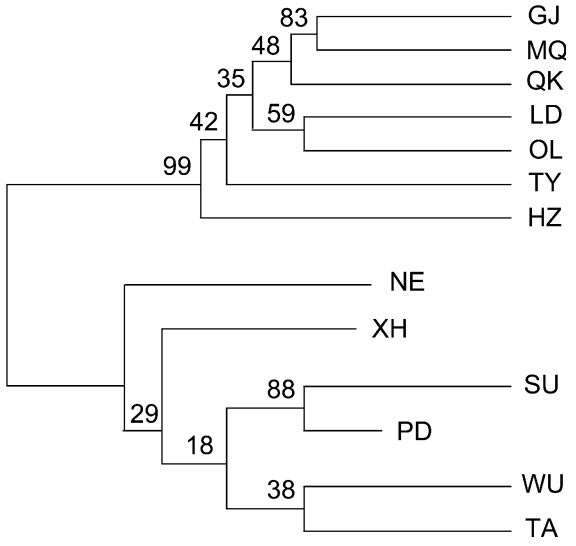

Figure 2. Neighbor-Joining (NJ) phylogeny based on Nei's genetic distance (D_A) for 11 indigenous Chinese sheep breeds, 1 improved breed and 1 Western breed (for breed abbreviations, see the legend for Figure 1). Bootstrap supports are shown, based on 1000 permutations.

Western breed. Based on the previous cytogenetic evidence that suggests a distinct pattern of the centromere index and relative length of chromosomes between Mongolian derived breeds and Tibetan derived breeds (Pang et al. 1998), TA was deemed to belong to the Mongolia group. Likewise, a previous RAPD analysis also provided evidence that TA had Mongolia origin (Gong et al. 2002). Therefore, several earlier genetic analyses indicate a closer relationships between TA and breeds of Mongolia origin, consistent with our results.

Not all genetic distances calculated for this study (e.g. between HZ and TY) are in agreement with current classification using morphological data, but the latter has been suggested to be insufficient for determining relationships between breeds (Arranz et al. 1998; Mburu et al. 2003). Artificial admixture of breeds through trade might be an explanation but waits to be verified. Tsunoda et al. (1999) classified Asian sheep into three groups: Indo-Pakistan, Tibetan and Mongolia (Tsunoda et al. 1999). Per Tsunoda's classification, the indigenous breeds in the study belong to the Mongolian or Tibetan groups, but the results from this study are not in good agreement with Tsunoda's classification. Based on several investigations (Zheng 1988; this study) and historical evidence (Sun et al. 2003b), it is believed that TY belongs to the Mongolian group and may also be related to Large-(Fat)-Tail Han. The ancestors of Large-(Fat)-Tail Han were perhaps the Fat-Tail sheep in Middle East and Near East and they were brought to Xinjiang, Gansu and Shaanxi provinces of China along the 'Silk Road' through trade and migrations. As documented in the book ''Sheep and Goat Breeds in China'' (Zheng 1988), a sheep breed was formed in the Han Dynasty (202 B.C.–A.D. 220), whose appearance and body structure were quite different from TY, Small-(Fat)-Tail Han or Large-(Fat)-Tail Han, but bore a strong resemblance to HZ inhabiting in the Qinba mountain regions of Shaanxi. The fact that HZ and TY of Mongolian origins were clustered together based on our data suggests that the two breeds had frequent genetic exchanges. However, disputes regarding whether HZ should be classified into the Mongolian or Tibetan group will continue and more data are needed to resolve this issue.

Caution should be taken in interpretation of the tests for neutrality. The fact that a locus does not conform to the assumption of neutrality does not necessarily implicate selection the sole evolutionary force, because substantial historical migrations can also lead to departures from neutrality. The outcome of neutrality tests can also be affected by selection favoring heterozygotes, leading to an equal frequency of alleles and thus a small F value. Conversely, selection for homozygotes can inflate the F value. Historical evidence indicates that both large-scale migrations and selection could have affected these breeds (Sun et al. 2003a), which is supported by our finding that over one third of markers significantly departed from neutrality. Although presumably neutral, microsatellite loci can be linked to functional genes.

In conclusion, the high polymorphism information contents at all 30 markers, which cover 19 chromosomes of sheep genome, imply substantial retention of natural variation from the source populations. Analysis of genetic differentiation revealed substantial differentiation among these breeds, and phylogenetic and phylogeographic analyses displayed a remarkable degree of consistency between geographic origins, breeding histories and the pattern of genetic differentiation.

Acknowledgements

We thank two anonymous reviewers for their comments on an early version of the manuscript. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 30371026, 30170515, 30370798, 30571034 and 30570424), the Public Basis Program for Natural Resources (Grant No. 2001DEA10006) and the Heilongjiang Province Department of Education Outstanding Overseas Scientist grant (Grant No. 1055HG009, to SR).

References

Arranz JJ, Bayon Y, San Primitivo F (1998) Genetic relationships among Spanish sheep using microsatellites. *Anim. Genet.*, **29**, 435–440.

Barker JSF (1994) A global protocol for determining genetic distance among domestic livestock breeds. *Proceedings of the 5th world congress on genetics applied to livestock production,* Guelph, Ontario, Canada, **21**, 501–508.

Bonnet E, Van de Peer Y (2002) zt: a software tool for simple and partial Mentel tests. *Journal of Statistical Software*, **7(10)**, 1–12. Available at http://www.psb.rug.ac.be/~erbon/mantel/.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, **32**, 314–331.

Chang H, Geng SM, Wu B, Li XY, Chen YC (1989) Efficiencies of different sampling strategies for estimation of gene frequencies in Chinese cattle breeds. *Acta Universitatis Septentrionali Occidentali Agriculture*, **17**, 30–37.

Chen YC (1995) The significance of conservation of domestic animal diversity in China *Biodiversity (China)*, **3**, 143–146.

Chen YC, Cao HH (2001) Diversity of Chinese yellow cattle breeds and their conservation. *Biodiversity (China)*, **9**, 275–283.

Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.*, **15**, 290–295.

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG, Beattie AE, Davies K, Maddox JF, Kappes SW, Stone RT, Nguyen TC, Penty JM, Lord EA, Broom JE, Buitkamp J, Schwaiger W, Epplen JT, Matthew P, Matthews ME, Hulme DJ, Beh KJ, McGraw RA, Beattie CW (1995) An autosomal genetic linkage map of the sheep genome. *Genetics*, **140**, 703–724.

Crawford AM, Montgomery GW, Pierson CA, Brown T, Dodds KG, Sunden S, Henry HM, Ede AJ, Swarbrick PA, Berryman T, Penty JM, Hill DF (1994) Sheep linkage mapping: Nineteen linkage groups derived from the analysis of paternal half-sib families. *Genetics*, **137**, 573–579.

de Gortari MJ, Freking BA, Kappes SM, Leymaster KA, Crawford AM, Stone RT, Beattie CW (1997) Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. *Anim. Genet.*, **28**, 274–290.

Fan B, Wang ZG, Li YJ, Zhao XL, Liu B, Zhao SH, Yu M, Li MH, Chen SL, Xiong TA, Li K (2002) Genetic variation analysis within and among Chinese indigenous swine populations using microsatellite markers. *Anim. Genet.*, **33**, 422–427.

Feng WQ, Ma YH, Chen YC (1997) Views on genetic resources of domestic animals in China. *Chin. J. Anim. Vet. Sci.*, **28**, 300–303.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Fulton JE, Delany ME (2003) Genetics. Poultry genetic resources – operation rescue needed. *Science*, **300**, 1667–1668.

Gong YF, Li XL, Liu ZZ, Li JQ (2002) Studies of random amplifiedpolymorphic DNA (RAPD) of main indigenous sheep breeds in China. *Hereditas (Beijing)*, **24**, 423–426.

Guti'errez-Espeleta GA, Kalinowski ST, Boyce WM, Hedrick PW (2000) Genetic variation and population structure in desert bighorn sheep: Implications for conservation. *Conserv. Genet.*, **1**, 3–15.

Hartl DL, Clark AG (1989) *Principles of Population Genetics*, Sinauer Associates, Inc., Sunderland, Massachusetts.

Li K, Chen Y, Moran C, Fan B, Zhao S, Peng Z (2000) Analysis of diversity and genetic relationships between four Chinese indigenous pig breeds and one Australian commercial pig breed. *Anim. Genet.*, **31**, 322–325.

Li MH, Zhao SH, Bian C, Wang HS, Wei H, Liu B, Yu M, Fan B, Chen SL, Zhu MJ, Li SJ, Xiong TA, Li K (2002) Genetic relationships among twelve Chinese indigenous goat populations based on microsatellite analysis. *Genet. Sel. Evol.*, **34**, 729–744.

Long JR, Qiu XP, Zeng FT, Tang LM, Zhang YP (2003) Origin of rabbit (Oryctolagus cuniculus) in China: Evidence from mitochondrial DNA control region sequence analysis. *Anim. Genet.*, **34**, 82–87.

Ma YH, Cao HH, Chen YC, Wang D, Lu SJ (2003a) Analysis of genetic diversity in several cattle breeds. *Agric. Sci. China*, **2**, 804–808.

Ma YH, Xu GF, Wang DY, Liu HL (2003b) Dynamic changes in animal genetic resources in China. *Agric. Sci. China*, **2**, 80–84.

Mburu DN, Ochieng JW, Kuria SG, Jianlin H, Kaufmann B, Rege JE, Hanotte O (2003) Genetic diversity and relationships of indigenous Kenyan camel (Camelus dromedarius) populations: Implications for their classification. *Anim. Genet.*, **34**, 26–32.

Nei M (1972) Genetic distance between population *Am. Nat.*, **106**, 283–293.

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

Ota T (1993) *Dispan: Genetic distance and phylogenetic analysis* Pennsylvania State University, University Park, PA. Available at: http://mep.bio.psu.edu/dispanreadme.html.

Pang YZ, Zhou YY, Xu TS, Zhang HW, Wang DY (1998) Karyotype analysis of Henan Small-Tail Han sheep. *Zhong Guo Xu Mu Zha Zhi (Chin. J. Anim. Sci.)*, **34**(2), 29–30.

Rao S, Notter DR (2000) Genetic analysis of litter size in Targhee, Suffolk, and Polypay sheep. *J. Anim. Sci.*, **78**, 2113–2120.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin: A software for population genetics data analysis. Ver 2.000* Genetics and Biometry Lab, Department of Anthropology, University of Geneva. Available at: http://lgb.unige.ch/arlequin/software/.

Simon DL (1999) Better decisions in conservation of farm animal genetic resources by use of international sources of information *Information in Decision Making in Agricultural Research and Practice/IAALD QB XLIV*, 1/2, 59–65.

Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*, Freeman, San Francisco.

Sun W, Chang H, Ren ZJ, Yang ZP, Geng RQ, Tsunoda K (2003a) Studies on discriminants phylogenetic extraction of sheep populations in East and South of Central Asia. *Acta Veterinaria et Zootechnica Sinica*, **34**, 132–138.

Sun W, Chang H, Yang ZP, Geng RQ, Lu SX, Chang GB, Xu W, Wang HY, Ni DX, Fan BS, Tsunoda K (2003b) Phylogenetic relationships between sheep populations in East and South of Central Asia. *Scientia Agriculture Sinica*, **36**, 94–98.

Tsunoda K, Nozawa K, Madeda Y, Tanabe Y (1999) External morphological characters and blood protein and non-protein polymorphism of Native sheep in Central Mongolia. *Rep. Soc. Res. Native Livestock*, **17**, 63–82.

Weir BS (1996) *Genetic Data Analysis 2: Methods for Discrete Population Genetic Data.* Sinauer Associates, Sunderland,

Massachusetts. Software GDA available at: http://lewis.eeb.uconn.edu/lewishome/software.html.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany*, **129**, 157. The corresponding software POPGENE available at: http://www.ualberta.ca/~fyeh/.

Zheng PL (1988) *Sheep and Goat Breeds in China*, Shanghai Scientific & Technology Publishers, Shanghai.