



Genetic variability and individual assignment of Chinese indigenous sheep populations (*Ovis aries*) using microsatellites

L. L. Niu*, H. B. Li*[†], Y. H. Ma[‡] and L. X. Du*

*National Centre for Molecular Genetics and Breeding of Animal, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China. [†]Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [‡]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China

Summary

The purpose of this study was to assess the genetic characteristics of six breeds of Chinese local sheep using 19 microsatellite loci and to effectively validate statistical methods for individual assignment based on informative microsatellites. All the six breeds deviated from Hardy–Weinberg equilibrium expectations, while the majority of markers complied. The polymorphism information content (PIC) of overall loci for the six populations ranged from 0.283 (*SRCRSP5*) to 0.852 (*OarVH72*). Tibetan sheep were the most diverse population with the highest mean allelic richness (6.895), while Ujmuqin (UQ) harboured the lowest allelic richness (6.000). The F -statistics for the six populations were $F_{IS} = -0.172$, $F_{IT} = -0.082$ and $F_{ST} = 0.077$, respectively. Furthermore, the pair-wise F_{IS} revealed a moderate genetic differentiation among populations ($P < 0.01$), indicating that all breeds can be considered genetically independent entities. The lowest genetic differentiation was between Tengchong (TC) and UQ ($F_{ST} = 0.041$), and the highest one was between TC and Fat-tailed Han ($F_{ST} = 0.111$). In comparing the three statistical models, we note that the seven microsatellite loci (*MAF65*, *OarJMP58*, *SRCRSP9*, *MCM140*, *OarAE129*, *BM8125* and *SRCRSP5*) commonly used for individual assignment will ensure a powerful detection of individual origin, with accuracy up to 91.87%, when the likelihood-based method is used. Overall, these findings shed light onto the genetic characteristics of Chinese indigenous sheep and offer a set of microsatellite loci that is simple, economic and highly informative for individual assignment of Chinese sheep.

Keywords Chinese local sheep, genetic characteristics, individual assignment, microsatellite.

As one of the earliest livestock species to be domesticated, sheep (*Ovis aries*) are distributed in broad areas, especially in China. Because of complicated topography, great gaps in altitude, and long distances between hinterland and ocean, China has diversified ecological conditions and climate types (North Temperate Zone and Subtropical Zone), which have profoundly influenced the formation and distribution of domestic animal diversity (National Consultative Committee 2004). It has been reported that there are 79 sheep breeds (31 indigenous, 9 cross, 10 introduced and some undocumented breeds) in China (Ma *et al.* 2002). During long-term adaptation to local environments, many specific merits have

accumulated in local sheep populations, such as high reproductive rates, high productivity, and resistance to disease. However, many breeds have been identified as being in danger of extinction because of blind crossbreeding that is being carried out without enough attention being given to conservation. For effective and accurate management of livestock breeds and assurance of food safety in both domestic and global trade, the traceability of individuals to their source breed is also becoming more and more important (McKean 2001). Microsatellite markers (MS) were identified as an effective tool in individual assignment and genetic variation surveys several decades ago (Tadano *et al.* 2008; Boitard *et al.* 2010; Spencer *et al.* 2010). Here, we describe the genetic variation of Chinese local sheep and identify an optimized set of microsatellite loci that may be used for individual assignment of Chinese sheep.

Unrelated sheep of both sexes were randomly collected from in situ preserved farms located in six geographically separated provinces in China (Fig. S1). A total of 237

Address for correspondence

L. X. Du, National Centre for Molecular Genetics and Breeding of Animal, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China.
E-mail: lxdu@263.net

Accepted for publication 2 February 2011

samples were collected from six local sheep populations: Fat-tailed Han (FTH), Tibetan (TB), Kazakh (HZK), Hu (HU), Ujmuqin (UQ) and TengChong (TC) (Table 1). Samples were selected according to tail phenotype, which was the main basis of breed classification in China (Zheng 1988) and trended to be less affected by artificial selection. FTH is a breed famous for lambskin and meat quality, especially with their long fat tails: the weight of the fat tail is around 10.5 kg. HZK has a long domestic history (Zheng 1988) and adapts well to pasture areas, resulting in its high meat production. TB and TC originate from Tibetan breeds and were domesticated separately by geographical isolation. HU is distributed south of the Yangtze River, is well adapted to the local damp climate, and is famous for the quality of its wool and strong ability to reproduce. Lastly, UQ is located in the north of China (Inner Mongolia) and is a breed raised for mutton. Ear tissue was collected and stored at -70°C in 75% ethanol. Genomic DNA was extracted following the phenol–chloroform method (Sambrook *et al.* 1989).

Nineteen MS from the panel recommended by FAO were used in this study (Table S1). Deviations from Hardy–Weinberg equilibrium (HWE) for each locus in each population were performed using GENEPop v.3.4 (Raymond & Rousset 1995). The polymorphism statistics were implemented in the EXCEL MICROSATELLITE TOOLKIT version 3.1.1 (Park 2001) and CERVUS 3.0 program (Kalinowski *et al.* 2007). Wright's F -statistics were calculated using FSTAT v.2.9.3 (Goudet 1995). A hierarchical analysis of variance was carried out using the molecular variance (AMOVA) implemented in the ARLEQUIN 3.5.1.2 package (Excoffier *et al.* 2005). Genetic distances were estimated by Dispan (Ota 1993) and MICROSAT v1.5d (Minch 1997). Individual assignment was estimated by STRUCTURE v2.2 (Pritchard *et al.* 2000), GENECLASS2 (Piry *et al.* 2004) and WHICHLOCI (Banks & Eichert 2000), respectively.

In general, the six breeds have a high genetic variation, which is consistent with previous studies (Sun *et al.* 2007; Zhong *et al.* 2010). All other loci displayed high-level polymorphisms, except *SRCRSP5* and *ILSTS11*, and the polymorphism information content (PIC) ranged from 0.283 (*SRCRSP5*) to 0.852 (*OarVH72*). Detailed information is listed in Tables 1 & S1. The majority of the markers did not show deviation from HWE, while the populations were a departure from HWE (suggesting unnatural mating among those populations; data were omitted).

F -statistics of overall loci are $F_{IS} = -0.172$, $F_{IT} = -0.082$, $F_{ST} = 0.077$ (Table S1), which indicate an excess of heterozygotes likely reflecting household-based management and moderate genetic differentiations. Both F -statistics and AMOVA analysis revealed a significant genetic differentiation among populations (Tables S1 and S2). In addition, distances data also indicate that the six sheep breeds can be considered to be genetically independent entities. The highest genetic differentiation was revealed

Table 1 The results of genetic statistical parameters and individual assignment based on different models.

Breed	No.	Morphology	GD	AR	H_0	H_e	F_{IS}	Percentage of individuals correctly assigned (%) ¹					
								Likelihood method			Distance method		
								Bayesian ²	ML ³	Bayesian ⁴	Frequency	DS	DA
Fat-tailed Han (FTH)	41	Long fat-tailed	0.711	6.421	0.709	0.711	0.004	100.00	97.60	100.00	100.00	100.00	
Kazakh (HZK)	42	Fat-rumped	0.714	6.737	0.868	0.716	-0.215	95.20	92.90	92.90	92.90	92.90	
Hu (HU)	41	Short fat-tailed	0.715	6.523	0.895	0.717	-0.252	85.40	95.10	92.70	95.10	95.10	
Tibetan (TB)	39	Short thin-tailed	0.729	6.895	0.846	0.730	-0.160	89.70	92.30	92.30	92.30	89.70	
Tengchong (TC)	38	Short thin-tailed	0.720	6.421	0.887	0.723	-0.232	73.70	81.60	81.60	81.60	78.90	
Ujmuqin (UQ)	36	Short fat-tailed	0.698	6.000	0.832	0.700	-0.191	86.10	91.70	91.70	88.90	91.70	
Mean			0.715	6.500	0.840	0.716	-0.174	88.35	91.87	91.87	91.80	91.83	

No., the number of animals tested per breed; GD, gene diversity; AR, allelic richness; H_0 , observed heterozygosity; H_e , expected heterozygosity; DS, Nei's standard genetic distance; DA, Nei's DA distance; DAS, shared allele distance.

¹Seven effective loci used for assignment, which were determined by WHICHLOCI program (Banks *et al.* 2003).

²Implemented in STRUCTURE.

³Maximum likelihood, implemented in WHICHRUN.

⁴Implemented in GENECLASS.

	FTH	HZK	HU	TB	TC	UQ
FTH		0.1898	0.1944	0.2763	0.2657	0.2691
HZK	0.0804		0.1275	0.1952	0.2260	0.2207
HU	0.0805	0.0552		0.1476	0.1569	0.1377
TB	0.1108	0.0758	0.0649		0.1137	0.1254
TC	0.1113	0.0958	0.0685	0.0449		0.0801
UQ	0.1090	0.0880	0.0598	0.0539	0.0407	

F_{ST} estimates are shown below the diagonal and DA matrix is above the diagonal. All estimates of F_{ST} were significant ($P < 0.01$).

FTH, Fat-tailed Han; HZK, Kazakh; TB, Tibetan; TC, TengChong; UQ, Ujmuqin.

between TC and FTH, while the lowest was between TC and UQ (Table 2). TC and FTH had different tail types and were separated by geography, which was consistent with the clustering results of Zhao *et al.* (2006). The most likely number of groups identified by STRUCTURE was $K = 5$. In view of breed phenotype (tail type), these results clearly presented the following groups: (i) FTH, long fat-tailed, (ii) HZK, fat rumped, (iii) TB, short thin-tailed, (iv) HU, short fat-tailed, and (v) TC and UQ. However, a discrepancy was revealed in the last group. TC (short thin-tailed) is a breed of Yunnan sheep without explicit domestic histories. Based on mtDNA analyses, three lineages have been detected in Yunnan sheep, as in Mongolian sheep (Guo *et al.* 2005; Chen *et al.* 2006). These studies indicated that the TC shares some genetic ancestry with UQ, which should also be revealed in Fig. S3. Furthermore, Zhong *et al.* (2010) also found obvious genetic differentiation between HU and UQ populations (representing fat-tailed type), which was consistent with our findings.

Seven loci (*MAF65*, *OarJMP58*, *SRCRSP9*, *MCM140*, *OarAE129*, *BM8125* and *SRCRSP5*) were identified by WHICHLOCI with 95.5% accuracy in assignment testing (Banks *et al.* 2003). The results of assignment testing are shown in Table 1. Maximum likelihood in WHICHRUN, Bayesian in GENECLASS, and frequency showed the same mean assignment correct percentage (91.87%). Three distance methods (DS, DA and DAS) obtained a mean 91.82% accuracy, and DA was slightly better than the others. Although STRUCTURE has been used to identify and assign subgroups in many other studies, the accuracy was only a maximum of 88.35% in our study. This may be explained by high genetic diversity included in the population data sets (Leroy *et al.* 2009). For assignment testing in Chinese local sheep, at least seven MS loci were confirmed effectively and powerfully. The six breeds represented the main morphology characteristics of the three indigenous sheep groups in China (Mongolian sheep, HZK sheep and TB sheep) (Zheng 1988). In general, our study was similar to previous work in goats (Fan *et al.* 2008) and beef cattle (Mukesh *et al.* 2009). Our study also reiterated the resolution of the Bayesian approach for correct assignment in comparison to genetic distance-based methods. Substantially high assignment accuracies for the FTH and HZK

Table 2 Pair-wise F_{ST} and Nei's genetic distances (DA) among the six local sheep breeds in China.

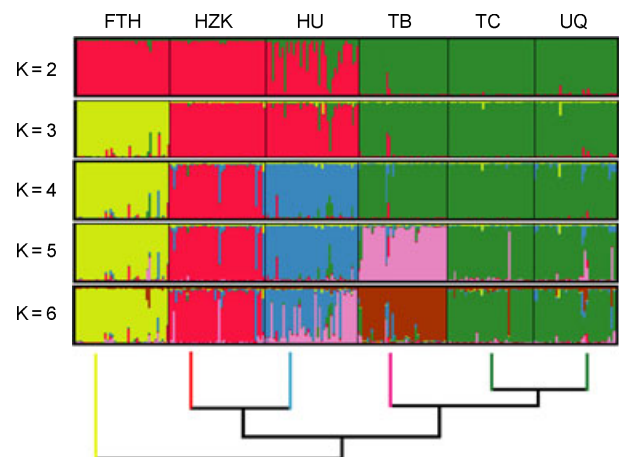


Figure 1 The genetic structure of the six local sheep breeds in China. A modified NJ tree is shown under the structure figure.

breeds indicate the relative isolation of genetic structure, while lower assignment rates for the TB, TC, HU and UQ breeds highlight the existence of some admixtures (Figs 1 & S2).

At present, the genetic diversity of livestock is facing a global 'livestock meltdown', as suggested by ILRI experts. In addition to genetic variation assessment, individual assignment testing is also an important part of the future conservation management of livestock breeds.

Acknowledgements

We thank the anonymous reviewers for their comments on an early version of this manuscript. This work was supported by the National Natural Resource Science and Technology Platform (No.2004DKA30450). We are grateful for the excellent technical support of Dr. Jun Guo. The authors also thank Dr. Tao Zhong for his helpful comments during the revision of this manuscript.

References

Banks M.A. & Eichert W. (2000) WHICHRUN (version 3.2): a computer program for population assignment of individuals

- based on multilocus genotype data. *Journal of Heredity* **91**, 87–9.
- Banks M.A., Eichert W. & Olsen J.B. (2003) Which genetic loci have greater population assignment power? *Bioinformatics* **19**, 1436–8.
- Boitard S., Chevalet C., Mercat M.J., Meriaux J.C., Sanchez A., Tibau J. & Sancristobal M. (2010) Genetic variability, structure and assignment of Spanish and French pig populations based on a large sampling. *Animal Genetics* **41**, 608–18.
- Chen S.Y., Duan Z.Y., Sha T., Xiangyu J., Wu S.F. & Zhang Y.P. (2006) Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene* **376**, 216–23.
- Excoffier L., Laval G. & Schneider S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50. (<http://cmpg.unibe.ch/software/arlequin3/>).
- Fan B., Han J.L., Chen S.L., Mburub D.N., Hanotte O., Chen Q.K., Zhao S.H. & Li K. (2008) Individual-breed assignments in caprine populations using microsatellite DNA analysis. *Small Ruminant Research* **75**, 154–61.
- Goudet J. (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86**, 485–6. (<http://www.unil.ch/izea/software/fstat.html>).
- Guo J., Du L.X., Ma Y.H., Guan W.J., Li H.B., Zhao Q.J., Li X. & Rao S.Q. (2005) A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics* **36**, 331–6.
- Kalinowski S.T., Taper M.L. & Marshall T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099–106.
- Leroy G., Verrier E., Meriaux J.C. & Rognon X. (2009) Genetic diversity of dog breeds: between-breed diversity, breed assignment and conservation approaches. *Animal Genetics* **40**, 333–43.
- Ma Y.H., Xu G.F., Wa D.Y. & Liu H.L. (2002) Study on dynamic information of animal genetic resources in China. *Scientia Agricultura Sinica* **35**, 552–5. (in Chinese).
- McKean J.D. (2001) The importance of traceability for public health and consumer protection. *Revue Scientifique et Technique (International Office of Epizootics)* **20**, 363–71.
- Minch E. (1997) MICROSAT, Version 1.5d: Stanford University Medical Center, Stanford. (<http://hpgl.stanford.edu/projects/microsat/>).
- Mukesh M., Sodhia M., Katariaa R.S. & Mishraa B.P. (2009) Use of microsatellite multilocus genotypic data for individual assignment assay in six native cattle breeds from north-western region of India. *Livestock Science* **121**, 72–7.
- National Consultative Committee (2004) *Country Report for the Preparation of SoW-AnGR Report on Domestic Animal Genetic Resources in China*. China Agriculture Press, Beijing, China.
- Ota T. (1993) Dispan: Genetic Distance and Phylogenetic Analysis. Pennsylvania State University, University Park. (<ftp://ftp.bio.indiana.edu/molbio/ibmpc/dispan.zip>).
- Park S.D.E. (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. Ph.D. thesis University of Dublin. (<http://animalgenomics.ucd.ie/sdepark/ms-toolkit/>).
- Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L. & Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536–9.
- Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59.
- Raymond M. & Rousset F. (1995) GENEPOP (version 3.4): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–9. (<http://kimura.univ-montp2.fr/~rousset/Genepop.htm>).
- Sambrook J., Fritsch E.F. & Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour, Cold Spring Harbour.
- Spencer P.B., Wilson K.J. & Tinson A. (2010) Parentage testing of racing camels (*Camelus dromedarius*) using microsatellite DNA typing. *Animal Genetics* **41**, 662–5.
- Sun W., Chang H., Ji D., Liao X., Du L., Lu S. & Kenji T. (2007) Analysis on genetic diversity and isolation mechanism by distance of different ecological type sheep breeds in Mongolia sheep group. *Journal of Genetics and Genomics* **34**, 1001–9.
- Tadano R., Nishibori M. & Tsudzuki M. (2008) High accuracy of genetic discrimination among chicken lines obtained through an individual assignment test. *Animal Genetics* **39**, 567–71.
- Zhao Q.J., Ma Y.H., Guan W.J., Chen H.Y., Yang J., Di R. & Su R. (2006) Microsatellite analysis of genetic diversity and phylogenetic relationship of 8 Tibetan sheep populations in China. *Animal Biotechnology Bulletin* **10**, 503–7. (in Chinese).
- Zheng P.L. (1988) *Sheep and Goat Breeds in China*. Shanghai Scientific & Technology Publishers, Shanghai. (in Chinese).
- Zhong T., Han J.L., Guo J., Zhao Q.J., Fu B.L., He X.H., Jeon J.T., Guan W.J. & Ma Y.H. (2010) Genetic diversity of Chinese indigenous sheep breeds inferred from microsatellite markers. *Small Ruminant Research* **90**, 88–94.
- Zhong T., Han J.L., Guo J., Zhao Q.J., Fu B.L., Pu Y.B., He X.H., Jeon J.T., Guan W.J. & Ma Y.H. (2011) Tracing genetic differentiation of Chinese Mongolian sheep using microsatellites. *Animal Genetics* doi: 10.1111/j.1365-2052.2011.02181.x.

Supporting information

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

Figure S1 The main tail-type distributions of Chinese indigenous sheep and the sampling locations in this study.

Figure S2 The NJ tree of the 237 individuals based on *Dps* distance matrix.

Figure S3 The NJ tree of the six sheep populations based on DA distances.

Table S1 Descriptive statistics of the 19 microsatellite loci.

Table S2 Partitioning of genetic variation by the Analysis of Molecular Variance (AMOVA).

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.