doi:10.1111/j.1365-2052.2005.01310.x

A novel maternal lineage revealed in sheep (*Ovis aries*)

J. Guo*,[1], L.-X. Du*,[1], Y.-H. Ma*,[1], W.-J. Guan*, H.-B. Li*, Q.-J. Zhao*, X. Li[†] and S.-Q. Rao[†,‡]

*Division of Animal Resources and Conservation, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100094, China. [†]Department of Bioinformatics, Harbin Medical University, Harbin 150086, China. [‡]Department of Molecular Cardiology and Cardiovascular Medicine, the Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA

Summary

It is generally believed that domestic sheep have two maternal lineages (haplotypes A and B), based on mitochondrial DNA analysis. In the present study, we provide evidence that a novel maternal lineage (haplotype C) is exhibited in Chinese native sheep. To verify this finding, 231 samples were collected from six Chinese local breeds, which cover the vast geographical region of sheep inhabitation in China. For comparison, 50 samples were collected from two Western breeds collected in China. Mitochondrial DNA was screened by PCR single-strand conformational polymorphism (SSCP), leading to the identification of novel band patterns in *ND2* and *ND4* genes in the Chinese breeds. Interestingly, mutations at the two loci were in strong linkage disequilibrium. Direct sequencing of the DNA fragments revealed a non-synonymous substitution in *ND2*. Furthermore, two synonymous mutations were identified by comparisons of the novel type (haplotype C) and the established types (haplotypes A and B). The entire mitochondrial control region for 55 samples was then sequenced to construct a phylogenetic tree and median joining network. Both the tree and network demonstrated a topology of three groups, which is in consistent with the SSCP analysis. Unlike Western breeds, Chinese breeds are composed mainly of haplotypes A and B, but with a small fraction of haplotype C. According to Fu's test and mismatch distribution, haplotype C has not been subject to a recent population expansion. Based on these results, we propose a novel origin for Chinese sheep.

Keywords maternal lineage, mitochondrial DNA, phylogenetics, sheep, the mitochondrial control region.

Introduction

The origin of domestic sheep (*Ovis aries*) remains uncertain and controversial. Based on archaeological evidence, sheep were probably first domesticated in the Fertile Crescent region of Southwest Asia 10 000 years ago (Ryder & Stephenson 1968). However, morphological evidence suggested that sheep might have been derived from urial (*Ovis vignei*), found mainly in the mountain ranges of Central Asia (Ryder & Stephenson 1968; Piper & Ruvinsky 1997). In fact, domestic sheep can interbreed with mouflon, urial and argali, which has imposed an additional complication for studying the origins of sheep.

Mitochondrial DNA (mtDNA) is an important material for phylogenetic inference. Animal mtDNA is deemed to strictly follow maternal inheritance and is highly variable within species. In particular, the control-region sequence (which includes the D loop) in the mitochondrial genome evolves very rapidly compared with nuclear DNA. Mitochondrial DNA can also tell recent demographic processes acting on a population. For example, analysis of mtDNA can reveal whether a population has undergone a recent demographic expansion or has a more complex history (Bruford *et al.* 2003).

Mitochondrial DNA has been widely used to explore the origins of sheep. According to previous phylogenetic trees constructed by mtDNA (Hiendleder *et al.* 1998), there are two haplotypes in domestic sheep. These maternal lineages correlate well, but not completely, with modern fat- and thin-tailed phenotypic varieties (Hiendleder *et al.* 1998; Wu

Address for correspondence

Ma, Division of Animal Resources and Conservation, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Yuanmingyuanxi Road, No. 2, Haidian District, Beijing 100094, China.
E-mail: yuehui.ma@iascaas.net.cn

Correspondence may also be addressed to Rao, Department of Molecular Cardiology and Cardiovascular Medicine, The Cleveland Clinic Foundation, Cleveland, OH 44195, USA. E-mail: raos@ccf.org

[1]These authors contributed equally to this work.

Accepted for publication 17 April 2005

et al. 2003). It is generally agreed that the most recent common ancestor of sheep is the mouflon (*Ovis musimom*). The hypothesis that argali (*Ovis ammon*) and urial are the putative ancestors has been rejected by mtDNA analysis (Hiendleder *et al.* 2002; Wu *et al.* 2003). However, the previous phylogenetic inference was based on data from Western breeds and may not extend to Eastern breeds. Data from East and South Asia have provided important additional information on domestication of several species such as dogs, swine and cattle, which were established in China and South Asia before or just after sheep were domesticated (Loftus *et al.* 1994; Giuffra *et al.* 2000; Leonard *et al.* 2002; Savolainen *et al.* 2002). Recently, some studies have suggested that an independent domestication in Pakistan gave rise to the Cashmere breeds (Meadow 1996; Porter & Tebbit 1996; Joshi *et al.* 2004). Therefore, data from Eastern countries such as China will expand the phylogenetic knowledge for sheep.

Materials and methods

Sample collection and DNA extraction

Our sampling covered six Chinese local breeds (Mongolian, Tibetan, Kazakh Fat-Ramped, Hu, Tong and Han) and two Western breeds (Polled Dorset and Texel) (see Table 1 and Fig. 1). Choices of breeds were based mainly on their distribution and tail types, including fat rump, short- and long-fat tail, short- and long-thin tail. Ear tissue for the investigated breeds was collected and stored at −70 °C in 75% ethanol before DNA extraction, described by Sambrook *et al.* (1989).

PCR and SSCP

For analysis of the conservative region of sheep mtDNA, eight pairs of primers were designed from known ovine sequences (GenBank accession no. NC_001941) using Primer 3.0 program (Primer Biosoft, Palo Alto, CA, USA). Primer OV6 was designed to amplify part of the *ND2* gene. The primer sequences were (i) 5′-CAACCCACGAGCCACAGAAG-3′ and (ii) 5′-CTGGGACTCAGAAGTGGAATGG-3′. The annealing temperature was 52 °C. Primer OV11 was designed to amplify part of gene *ND4* and its sequences were (i) 5′-GACTCCACCTCTGACTTCC-3′ and (ii) 5′-TGAATGAGAATGGCAACA-3′. The annealing temperature was 54 °C. The PCR amplification was carried out in a 12-μl reaction volume containing 20–100 ng DNA template, 1.0 μM of each primer, 200 μmol of dNTPs, 1x PCR buffer, 1.5 mM $MgCl_2$, and one unit of *Taq* DNA polymerase (TianWei Co., Shanghai, China). Amplification was performed in a PTC200^{TM} Programmable Thermal Controller (MJ Research® Inc., Waltham, MA, USA). In the experiment, 12 μl aliquots of the amplified samples were mixed with 12 μl of formamide loading dye (98% formamide, 10 mM $EDTANa_2$, 0.02% xylene cyanol, 0.02% bromophenol blue and add double distilled water to 100%). The samples were subsequently denatured by heating at 98 °C for 10 min, and were then placed on ice. The samples were loaded onto 12% non-denaturing PAA gel (PAA acrylamide:bis acrylamide, 39:1) in TBE buffer (0.89 M tris–base, 0.89 M boric acid, 20 mM EDTA, pH 8.0). Vertical electrophoresis was performed at 4 W in an icebox at 4 °C for 10–16 h, using the PAC 3000 power system (Bio-Rad, Hercules, CA, USA). The PCR products, randomly selected from each lineage, were cloned into *pGEM*–T easy vector (Promega Inc., Madison, WI, USA) according to the manufacturer's instructions. Positive clones were bidirectionally sequenced by BioAsian Co. (Beijing, China).

Sequencing the mitochondrial control region

Primers for the entire ovine mitochondrial control region were described previously (Hiendleder *et al.* 2002). The sequences were as follows: forward 5′-TCATCTAGGCATTTTCAGTG-3′ and reverse 5′-CTCACCATCAACCCCCAAAGC-3′. Amplification of PCR was carried out in a 60-μl reaction volume. The PCR conditions were an initial denaturation at 94 °C for 4 min, followed by 30 cycles at 94 °C for 30 s, 50 °C for 30 s, 72 °C for 90 s and a final extension of 8 min at 72 °C. Amplified DNA fragments were subject to electrophoresis on a 1% agarose gel and then purified with a column kit (TianWei Co.). The remaining

Table 1 Characteristics of eight sheep breeds.

Breeds	Sample size	Characteristics
Mongolian	26	Short-fat tail; rams have horn and ewes are polled; white fleece colour, sometimes with black or brown face
Tibetan	46	Short thin tail; an ancient breed, adapted to the unfavourable environments of the Tibetan plateau
Kazakh Fat-Rumped	48	Fat rump; ewes are polled or have a tiny horn; brown fleece colour, sometimes, with yellow head and limbs
Hu	32	Short-fat tail; polled; prolific
Tong	41	Long-fat tail; a tiny horn
Han	38	Short-fat tail; prolific; rams have horn and ewes are polled
Polled Dorset	30	Long-thin tail; polled; white fleece colour
Texel	20	Long-thin tail; polled; usually white fleece, sometimes mixed with blond, occasionally with black spots on nose and legs

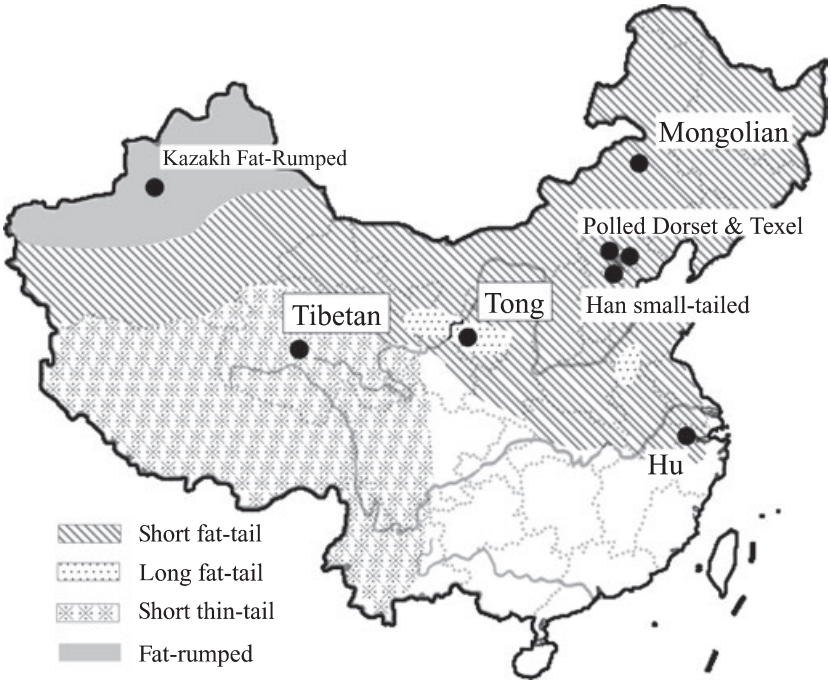

Figure 1 Geographical distributions of sheep breeds included in the study. Samples of two Western breeds, which were imported into China in 1997, were collected from Beijing.

steps for sequencing (cloning and sequencing) were carried out as performed in the single-strand conformational polymorphism (SSCP) analysis, but only one strand was sequenced.

Data analysis

Sequence from the entire mtDNA control region of 55 individuals from eight sheep breeds was aligned using the Clustal X 1.83 package (Thompson *et al.* 1994). The tandem sequence(s) in the control region in all samples was omitted because the accumulation of mutations in the repeated region has occurred at different rates. The neighbour-joining method (NJ) or unweighted pair-group method with arithmetic averaging (UPGMA) tree was constructed using the program Mega 2.0 (Kumar *et al.* 2001), with a Kimura 2-parameter (transition only) model and a bootstrap (number of replications = 1000) test. A maximum-likelihood (ML) tree was constructed with Tree-Puzzle 5.0 (Schmidt *et al.* 2002). Rooting was determined with a goat control region outgroup and other parameters were set at the default values. The median joining network was drawn using the program Network 4.1.0.9 (Bandelt *et al.* 1999). Fu's Fs statistics (Fu 1997) were computed using Arlequin (http://lgb.unige.ch/arlequin/software/).

Results

SSCP analysis

We screened 281 mtDNA samples of six Chinese local breeds and two Western breeds. Two (OV6 and OV11) of eight pairs of primers revealed SSCP patterns not previously reported. For OV6, 12 PCR products were sequenced and three haplotypes of A ($n = 3$), B ($n = 3$) and novel type C ($n = 6$) were identified. Three haplotypes were also identified with OV11: A ($n = 6$), B ($n = 4$) and novel type C ($n = 3$). Based on the sequence data (GenBank accession nos AY827572–AY827575), we identified a non-synonymous mutation (4208 C/T) and a synonymous mutation (42224 A/G) in *ND2* and a synonymous mutation (10924 A/G) in *ND4*.

In a previous study, Hiendleder and co-workers identified haplotypes A and B that were used to distinguish the origins of Asian and Western sheep (Hiendleder 1998). The new haplotype C is unique for Chinese breeds and did not appear in the two Western breeds (Texel and Poll Dorset) examined in our study. Distributions of the three haplotypes in eight breeds are shown in Table 2.

Analysis of the mitochondrial control region

In addition to the SSCP analysis, the entire mtDNA control region of 55 individuals from eight breeds was sequenced. The corresponding sampling scheme was designed to include at least one sample for each haplotype per breed. A phylogenetic tree was constructed by NJ, UPGMA or ML. All the phylogenetic analysis methods identified the same three clades, reconciled with haplotypes A, B and C. Therefore, we show only the NJ tree constructed by Mega 2.0 (Fig. 2).

We also performed a network analysis to define the genetic structure between the lineages. The three clusters corresponded to the phylogenetic clades (Fig. 3). Because the three clades were clearly isolated, we believe that the evolution of each lineage was an independent event. It is

Table 2 Mitochondrial haplotype[1] frequencies in eight sheep breeds.

Breed	Frequency (%)		
	A (Asian type)	B (European type)	C (Novel type)
Mongolian	57.7 (15/26)	30.8 (8/26)	11.5 (3/26)
Kazakh Fat-Rumped	68.8 (33/48)	22.9 (11/48)	8.3 (4/48)
Tibetan	82.6 (38/46)	8.7 (4/46)	8.7 (4/46)
Tong	70.7 (29/41)	26.8 (11/41)	2.4 (1/41)
Han (small-tailed)	55.3 (21/38)	28.9 (11/38)	15.8 (6/38)
Hu	62.5 (20/32)	21.9 (7/32)	15.6 (5/32)
Polled Dorset	30.0 (9/30)	70.0 (21/30)	0.0
Texel	25.0 (5/20)	75.0 (15/20)	0.0

[1]Definitions of haplotyes A and B were based on the previously reported PCR-RFLP results (Hiendleder *et al.* 1998). In this study, genotyping was performed using single-strand conformational polymorphism.

noteworthy that several sequences, including 1 Mongolian (M1516), 1 Hu (L1530) and 1 Kazakh Fat-Rumped sheep (H1540) deviated from the clusters.

A sudden burst of population growth can impose an effect upon the distribution of mutations among homologous DNA sequences, which can be revealed by looking at the histograms of the number of nucleotide mismatches among all pair-wise comparisons – known as the mismatch distribution. For non-recombinant DNA sequences such as mtDNA, after a burst of growth, the mismatch distribution approaches a smooth curve with a single and well-defined peak, whereas in a steady-state population with a constant population size, the histogram is ragged (Rogers & Harpending 1992; Harpending 1994; Harpending *et al.* 1998; Hartl 2004). The inference of a rapid and recent expansion of haplotypes A and B can be made by studying the sequences of the control region. Mismatch distribution analysis revealed the genetic signatures for population expansion (data not shown), which can separate the different expansion pattern for the three sheep mtDNA lineages. The bell-shaped mismatch distributions of haplotypes A and B are consistent with a demographic population expansion. In contrast, the distribution of lineage C was ragged, suggesting that haplotype C had not gone through an recent expansion. An alternative test for population expansion is Fu's Fs statistics (Fu 1997), which gives a significant negative value when a population expansion occurs. Based on the results from Arlequin, we obtained the probability of observing a random neutral sample with a number of alleles similar or smaller than the observed value. The Fs values for haplotypes A, B and C were −8.20 ($P = 0.003$), −10.88 ($P < 0.001$) and −3.02 ($P = 0.070$), respectively, suggesting that haplotypes A and B departed significantly from the neutral model. Based on analyses of the Fu's Fs statistic and mismatch distributions, we concluded that haplotype C had a quite different demographic history from the other two lineages.

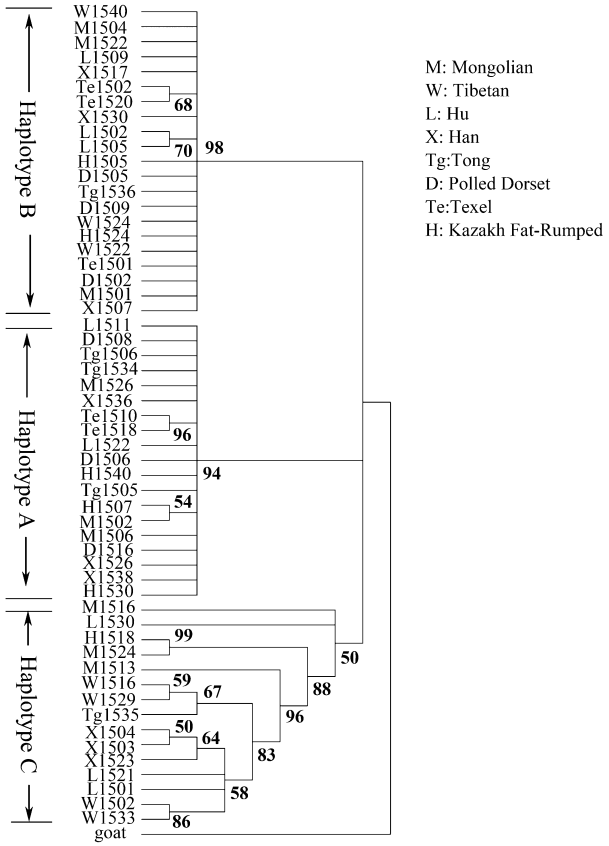

Figure 2 Neighbour-joining tree constructed by using the data from six Chinese local breeds and two Western breeds. The numbers at branches are bootstrap values obtained from 1000 simulated replicates (the values <50% were not shown). To achieve clarity, genetic distances were not scaled proportionally according to the magnitudes of the estimates. The names of breeds were based on DAD-IS (FAO) nomenclature system.

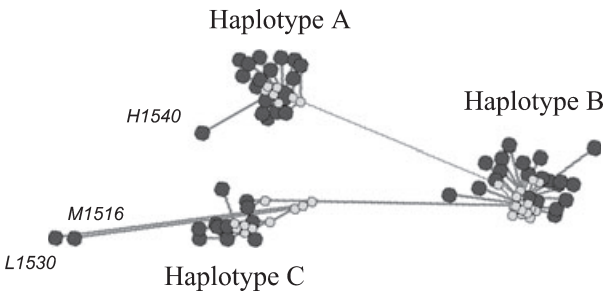

Figure 3 Network of three lineages in *Ovis aries*. The dark circles are 55 samples and the empty circles are median vectors.

Discussion

In this study, we conducted a comprehensive phylogenetic analysis of Chinese local breeds to obtain information about the genetic origin(s) of sheep that inhabit diverse regions of China. We provide convincing evidence to support the

presence of a novel mitochondrial haplotype in Chinese local breeds. It is evident from the analysis of mitochondrial genome that Chinese local breeds may have more involved phylogenetic history than Western breeds. The later were comprised chiefly of haplotype B (Hiendleder *et al.* 1998), but totally devoid of haplotype C. Although Chinese local sheep contain high frequencies of haplotypes A and B, haplotype C was identified in all of six Chinese local sheep breeds, with various frequencies. Two different approaches (SSCP mutation assay and the control region sequence analysis) have been used to verify the phylogenetic hypothesis.

According to mismatch distributions and Fu's statistics, lineage C did not experience a recent population expansion, which is different from other lineages identified in sheep or other farm animals. Unlike lineages A and B, lineage C had more complex phylogenetic branches, indicating that lineage C might be derived from a number of founders instead of a single common ancestor. Therefore, although lineage C might have undergone a historical expansion, the number of nucleotide mismatches was not changed sharply, leading to a non-bell-shaped mismatch distribution. An alternative explanation is that the mutation(s) in lineage C is(are) slightly deleterious to population viability and fecundity. There is increasing evidence for association between mtDNA mutations and male fertility (Gemmell *et al.* 2004). Studies on human, primate species, insects and birds suggest that mtDNA might underlie the phenotypic differences observed in sperm (Moore & Reijo-Pera 2000; Ward 2000; St John *et al.* 2001; Anderson & Dixson 2002; Froman *et al.* 2002).

Using a calibration of 5–7 Myr for the sheep–goat split derived from the fossil record (Savage 1983; Carroll 1987), the mutation rate can be estimated as $5.2–6.2 \times 10^{-5}$. The unweighted mean pair-wise differences of lineages A and B were 10.4 and 13.9 respectively. Thus, the time for lineage expansion might be 84 000–100 000 and 112 000–134 000 years ago for lineages A and B respectively. Unfortunately, these estimates do not agree with the dates of sheep domestication. A possible explanation is that the mismatches were overestimated.

In conclusion, we discovered a novel maternal lineage in Chinese local sheep breeds. This finding is useful to resolve the issue for the origin(s) of sheep. However, a wider phylogenetic inference of origin(s) of sheep in the Eastern waits for more data from the ancient domestication centres of Iran, Iraq, Jordan, Syria and Turkey.

Acknowledgements

We thank two anonymous referees for helpful comments on the manuscript. This project was supported in part by the National Natural Science Foundation of China (grant nos 30371026, 30170515 and 30370798), the Public Basis Programme for Natural Resources (grant no. 2001DEA10006) and the Heilongjiang Province Department of Education Outstanding Overseas Scientist grant (to SR).

References

Anderson M.J. & Dixson A.F. (2002) Sperm competition: motility and the midpiece in primates. *Nature* **416,** 496.

Bandelt H.J., Forster P. & Rohl A. (1999) Median-joining networks for inferring intraspecific phylogenies [network 4.x is available from http://www.fluxus-engineering.com/sharenet.htm]. *Molecular Biology and Evolution* **16,** 37–48.

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

Carroll R.L. (1987) *Vertebrate Paleontology and Evolution.* W.H. Freeman & Company, New York, NY.

Froman D.P., Pizzari T., Feltmann A.J., Castillo-Juarez H. & Birkhead T.R. (2002) Sperm mobility: mechanisms of fertilizing efficiency, genetic variation and phenotypic relationship with male status in the domestic fowl, *Gallus gallus domesticus*. *Proceedings of the Royal Society of London Series B Biological Science* **269,** 607–12.

Fu Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–25.

Gemmell N.J., Metcalf V.J. & Allendorf F.W. (2004) Mother's curse – the effect of mtDNA on population viability. *Trends in Ecology and Evolution* **19,** 238–44.

Giuffra E., Kijas J.M., Amarger V., Carlborg O., Jeon J.T. & Andersson L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154,** 1785–91.

Harpending H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66,** 591–600.

Harpending H.C., Batzer M.A., Gurven M., Jorde L.B., Rogers A.R. & Sherry S.T. (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 1961–7.

Hartl D.L. (2004) The origin of malaria: mixed messages from genetic diversity. *Nature Reviews Microbiology* **2,** 15–22.

Hiendleder S. (1998) A low rate of replacement substitutions in two major *Ovis aries* mitochondrial genomes. *Animal Genetics* **29,** 116–22.

Hiendleder S., Mainz K., Plante Y. & Lewalski H. (1998) Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *Journal of Heredity* **89,** 113–20.

Hiendleder S., Kaupe B., Wassmuth R. & Janke A. (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London Series B Biological Science* **269,** 893–904.

Joshi M.B., Rout P.K., Mandal A.K., Tyler-Smith C., Singh L. & Thangaraj K. (2004) Phylogeography and origin of Indian domestic goats. *Molecular Biology and Evolution* **21,** 454–62.

Kumar S., Tamura K., Jakobsen I.B. & Nei M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17,** 1244–5.

Leonard J.A., Wayne R.K., Wheeler J., Valadez R., Guillen S. & Vila C. (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* **298,** 1613–6.

Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 2757–61.

Meadow A. (1996) Animal domestication in the Middle East: a revised view from the eastern margin. In: *Harappan Civilization* (Ed. by G.L. Possehl), pp. 295–320. Oxford and IBH, New Delhi, India.

Moore F.L. & Reijo-Pera R.A. (2000) Male sperm motility dictated by mother's mtDNA. *American Journal of Human Genetics* **67,** 543–8.

Piper L. & Ruvinsky A. (1997) *The Genetics of Sheep*. CABI Publishing, Wallingford, UK.

Porter V. & Tebbit J. (1996) *Goats of the World*. Farming Press Limited, Ipswich, UK.

Rogers A.R. & Harpending H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9,** 552–69.

Ryder M.L. & Stephenson S.K. (1968) *Wool Growth*. Academic Press, San Diego, CA.

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

Savage D.E. (1983) *Mammalian Paleofaunas of the World*. Addison-Wesley, Reading, MA.

Savolainen P., Zhang Y.P., Luo J., Lundeberg J. & Leitner T. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* **298,** 1610–3.

Schmidt H.A., Strimmer K., Vingron M. & von Haeseler A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502–4.

St John J.C., Jokhi R.P. & Barratt C.L. (2001) Men with oligoasthenoteratozoospermia harbour higher numbers of multiple mitochondrial DNA deletions in their spermatozoa, but individual deletions are not indicative of overall aetiology. *Molecular Human Reproduction* **7,** 103–11.

Thompson J.D., Higgins D.G. & Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22,** 4673–80.

Ward P. (2000) Sperm length is heritable and sex-linked in the yellow dung fly (*Scatophaga stercoraria*). *Journal of Zoology* **251,** 349–53.

Wu C.H., Zhang Y.P., Bunch T.D., Wang S. & Wang W. (2003) Mitochondrial control region sequence variation within the argali wild sheep (*Ovis ammon*): evolution and conservation relevance. *Mammalia* **67,** 109–18.

Data deposition: the sequences reported in this paper have been deposited in the GenBank database (accession nos AY827572–AY827575).