DOI: 10.1534/genetics.106.068353

Five Ovine Mitochondrial Lineages Identified From Sheep Breeds of the Near East

Jennifer R. S. Meadows,*,† Ibrahim Cemal,‡ Orhan Karaca,‡ Elisha Gootwine§ and James W. Kijas*,[1]

**CSIRO Livestock Industries, St. Lucia, Brisbane 4067, Queensland, Australia, †University of New England, School of Rural Science and Agriculture, Armidale 2351, Australia, ‡Faculty of Agriculture, Department of Animal Science, Adnan Menderes University, Aydin 09100, Turkey and §Institute of Animal Science, Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel*

Manuscript received November 15, 2006
Accepted for publication December 21, 2006

ABSTRACT

Archaeozoological evidence indicates that sheep were first domesticated in the Fertile Crescent. To search for DNA sequence diversity arising from previously undetected domestication events, this survey examined nine breeds of sheep from modern-day Turkey and Israel. A total of 2027 bp of mitochondrial DNA (mtDNA) sequence from 197 sheep revealed a total of 85 haplotypes and a high level of genetic diversity. Six individuals carried three haplotypes, which clustered separately from the known ovine mtDNA lineages A, B, and C. Analysis of genetic distance, mismatch distribution, and comparisons with wild sheep confirmed that these represent two additional mtDNA lineages denoted D and E. The two haplogroup E sequences were found to link the previously identified groups A and C. The single haplogroup D sequence branched with the eastern mouflon (*Ovis orientalis*), urial (*O. vignei*), and argali (*O. ammon*) sheep. High sequence diversity ($K = 1.86\%$, haplogroup D and *O. orientalis*) indicates that the wild progenitor of this domestic lineage remains unresolved. The identification in this study of evidence for additional domestication events adds to the emerging view that sheep were recruited from wild populations multiple times in the same way as for other livestock species such as goat, cattle, and pig.

ARCHAEOZOOLOGICAL evidence from the ancient Levant points to the Pre-Pottery Neolithic B period, 9000–8000 years ago, as the time when sheep were first herded from the wild, tamed, and domesticated (reviewed by Legge 1996). The form of this wild ancestral population and the number of times and the process of its domestication remain unknown, as does its genetic contribution to the >1400 breeds (Scherf 2000) currently recognized in today's agricultural systems.

Mitochondrial sequencing has been used to elucidate the complexity and origins of many modern domestic livestock species, leading to a general theme of multiple maternal lineages. Recent studies of pigs (Larson *et al.* 2005) and goats (Joshi *et al.* 2004; Sardina *et al.* 2006), both thought to have origins in the Fertile Crescent, have revealed additional maternal clades. Following a wider geographic sampling of animals, goats now have six recognized lineages in Europe and Asia, while the six pig lineages span Europe, Asia, and the Pacific. Is this, then, a case of the more regions sampled, the more lineages will be found?

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ851886–DQ8552279.

[1]*Corresponding author:* CSIRO Livestock Industries, 306 Carmody Rd., St. Lucia, Brisbane 4067, Queensland, Australia.
E-mail: james.kijas@csiro.au

In 1996, Wood and Phua identified two domestic ovine lineages in sheep from New Zealand, and in 1998, Hiendleder *et al.* characterized these as of Asian (clade A) or European (clade B) origin after comparing the distribution of haplotypes within multiple breeds sampled from Germany, Russia, and Kazakhstan. The expansion to three recognized clades occurred in 2005 when Guo *et al.* and Pedrosa *et al.* sampled local breeds from China and Turkey, respectively. Clade C sequences have been reported at low frequency in sheep native to Portugal (Pereira *et al.* 2006), leading to an hypothesized gene flow from the Fertile Crescent to the Iberian Peninsula. Clade C has also been shown to contain more genetic diversity than either A or B (Pedrosa *et al.* 2005), but unlike clade B, and more in keeping with clade A, does not align with any wild Ovis animals. The European mouflon, *Ovis musimon*, is aligned to clade B (Hiendleder *et al.* 2002). Most recently, a single Karachai animal sampled from the north Caucasus revealed control region sequence, which grouped separately from the three defined ovine mitochondrial DNA (mtDNA) clades (Tapio *et al.* 2006). This was taken as evidence for a fourth maternal lineage and termed group D.

Ovine mitochondrial clade structure and global distribution patterns have been examined using network diagrams generated from collated published control region sequence (Chen *et al.* 2006; Pereira *et al.* 2006).

These phylogenies, based on a maximum of 531 bp, show clade B to be dominated by animals localized to Europe, clade C by sheep from the Middle East and Asia, and clade A to be a mixture from the Middle East, Asia, and Europe. Generally, these three lineages form starburst clusters, evidence of population expansion, but distant haplotypes are apparent in each group. In clade A, these outliers have been used to suggest group substructure and perhaps a more complicated ovine population history (Chen *et al.* 2006).

The aim of this study was to sample multiple breeds domesticated in the Near East (Turkey and Israel), to search for additional maternal lineages, and to confirm the genetic diversity previously reported in this region, the center of the Neolithic agricultural revolution (reviewed by Legge 1996). Two segments of the mitochondria, the control region and cytochrome *B*, were assayed to provide additional sequence and definition to identified phylogenies and to allow clade expansion-time estimates to be calculated.

MATERIALS AND METHODS

Animal resources: DNA was obtained from 197 unrelated animals representing eight breeds from Turkey and one from Israel. From eastern Turkey, two varieties of Akkaraman, the Karakas (KK, $n = 20$) and the Norduz (NZ, $n = 15$), were sampled from the province of Van, while the Morkaraman (MK, $n = 19$) and Tuj (TJ, $n = 16$) were collected from Erzurum. The western province of Aydin provided the Cine Capari (CC, $n = 14$), Sakiz (SZ, $n = 17$), and Karya (a Sakiz and Kivircik cross, KR, $n = 24$) samples, and the Karayaka (KY, $n = 15$) were sourced from the northern provinces of Samsun and Tokat. Cine Capari and Karayaka animals were selected from three and two flocks, respectively, while the remaining animals were sampled from within single flocks. Pedigree information, where available, was used to select unrelated animals; however, those taken from within the same flock should be considered related. Improved Awassi (AW; $n = 57$) sheep were systematically selected from the Kibbutz Ein Harod Ihud flock (Bet Shean Valley, Israel) using stud records to ensure their unrelated status for at least four generations maternally. DNA for KK, NZ, and AW were prepared from whole blood using the QIAamp DNA mini kit (QIAGEN, Doncaster, VIC, Australia) following the manufacturer's instructions. All other samples were extracted using a DNA salting protocol described elsewhere (Miller *et al.* 1988).

Mitochondrial DNA sequencing: Three fragments of the mitochondrion (mt) were generated using primers designed from the complete ovine mtDNA (AF010406). A 1246-bp fragment (mtCR) encompassing part of the control region, tRNA-Phe, and 12s rRNA (AF010406 positions 15,983–592), and a 1272-bp fragment (*cytB*) of the cytochrome *B* gene (AF010406 positions 14,078–15,349) were amplified using primer pairs mtcrF2/mtcrR1 and cytbF/cytbR as described by Meadows *et al.* (2005). PCR products from 197 individuals (Table 1) were directly sequenced in both the forward and reverse direction using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA) chemistry and visualized with a 3130xl Genetic Analyzer (Applied Biosystems). Sequence reads were aligned with Sequencher 4.2.2 (Gene Codes, Ann Arbor, MI) and trimmed to 1060 bp (mtCR) and 967 bp (*cytB*) as described by Meadows *et al.* (2005). The third fragment, a 721-bp region (hv-mtCR) of the highly variable control region (AF010406 positions 15,541–16,261), was collected using the primers and methods described by Tapio *et al.* (2006). This fragment was amplified and sequenced as above, using two individuals selected to represent each of the haplogroups defined in this study (HA–HE, $n = 10$). To allow direct comparison to the *cytB* and hv-mtCR fragments generated here, the following wild and domestic sequences were obtained from GenBank. For *cytB*, two domestic and three wild sequences were obtained: a representative of clade C (DQ097429), a divergent clade C haplotype (DQ097430), *O. vignei* (AF034729), *O. orientalis* (AJ867261), and *O. ammon* (AJ867272). For hv-mtCR, the single group D sequence described by Tapio *et al.* (2006) (DQ242212) and a divergent haplotype found in a Mongolian animal (AY829402) were acquired.

Sequence comparison and phylogenetic inference: The aligned sequence was assembled into three data sets. Analysis of cytochrome *B* gene sequences (*cytB*, 967 bp; $n = 202$) was used for comparison between domestic and wild sheep. Analysis of the larger combined mtCR-*cytB* data set (2027 bp; $n = 197$) was used to explore haplotype relationships between domestic sheep, and the third sequence set (hv-mtCR, 422 bp; $n = 12$) was used to facilitate direct comparison between divergent haplotypes identified here and by others (Guo *et al.* 2005; Tapio *et al.* 2006). The appropriate substitution model for the *cytB* and mtCR-*cytB* data sets was determined using hierarchical likelihood-ratio tests using MODELTEST 3.0.6 6 (Posada and Crandall 1998) as implemented in PAUP* 4.0, Macintosh Beta v10 (Swofford 2003). The Hasegawa–Kishino–Yano (HKY) (Hasegawa *et al.* 1985) evolutionary model with gamma distribution ($\Gamma = 0.251$) and HKY with invariable sites ($I = 0.844$) and gamma distribution ($\Gamma = 0.900$) were identified as the best models for the *cytB* and the mtCR-*cytB* data sets, respectively. Phylogenetic reconstruction was performed using multiple methods. MEGA 3.1 (Kumar *et al.* 2004) was used for the construction of bootstrap (1000 replications)-supported neighbor-joining (nj) trees, but as the software does not support HKY, the more inclusive Tamura–Nei (Tamura and Nei 1993) model was used for the estimation of genetic distance and tree construction. A Bayesian-derived consensus tree was constructed using the HKY + I + Γ model in MR BAYES 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). The software default priors were assumed and run for six million iterations to ensure the MCMC analysis reached convergence.

Indices of total sequence variation, nucleotide diversity (π), number of nucleotide differences (D), number of nucleotide substitutions per site (K), and haplotype structure were calculated with DnaSP 4.0 (Rozas *et al.* 2003). The relationships between haplotypes in the mtCR-*cytB* data set were visualized as a conservative ($\varepsilon = 0$) median-joining diagram (Bandelt *et al.* 1999) constructed using Network 4.1.1.2 (http://www.fluxus-engineering.com). Nucleotide weighting (w) was adjusted to reflect the difference in mutational frequency among indels ($w = 30$), transversions ($w = 20$), and transitions ($w = 10$), where the least-common event received the highest value.

Analysis of population expansion: The number of observed haplotypes was compared with the observed number of pairwise differences to search for signatures of population expansion using Fu's F_s statistic (Fu 1997) as calculated in Arlequin 3.01 (Excoffier *et al.* 2005). Observed mismatch distributions within haplogroups A, B, and C were then fitted against model parameters, assuming a sudden expansion model (Rogers and Harpending 1992; Rogers 1995) and a *P*-value test reported as described by Schneider and Excoffier (1999). For those sequence sets fitting the sudden expansion model, the time in years since commencement of the expansion

(t) was estimated from $t = \tau/2u$ using a nonlinear-stepwise least-squares approach in Arlequin 3.01 (Excoffier *et al.* 2005) and the 95% confidence interval was provided by 10,000 parametric bootstrap replicates. Tau (τ) is the empirical peak of the mismatch distribution and $u = m_t\mu$, where m_t is the length of sequence (967 bp) and μ is the substitution rate. An Ovis specific substitution rate, $\mu_{\text{Ovis}} = 2.51 \times 10^{-8}$ substitutions/site/year, was used, which is within the range previously reported for mammalian cytochrome *B* sequence (Pesole *et al.* 1999). μ_{Ovis} was calculated from estimates of lineage-specific sequence divergence (K_{Ovis}) and time of divergence (TOD_{Ovis}) using the relationship described by Li (1997). The TOD_{Ovis} was set to 473,000 years for the most recent common ancestor of the Moufloniform (incorporating *O. aries*) and Argaliform II (*O. ammon*) split as reported by Bunch *et al.* (2006). K_{Ovis} (0.0238 substitutions/site) was calculated as the net between-group genetic distance using MEGA 3.1 (Kumar *et al.* 2004) where the groups compared were 13 Argaliform II (*O. ammon*) *cytB* sequences (Bunch *et al.* 2006) and 197 domestic sequences (this study).

RESULTS

Sequence variation and genetic diversity: Two fragments of the ovine mitochondrion, mtCR (control region, 525 bp; tRNA-Phe and 12s rRNA, 535 bp) and *cytB* (cytochrome *B* gene, 967 bp) were sequenced for 197 animals of Turkish and Israeli origin (DQ851886–DQ8552279). Sequence alignment revealed the presence of an indel and 115 variable nucleotide positions, 77 of which proved to be phylogenetically informative. Mitochondrial nucleotide diversity ($\pi = 5.73 \pm 0.33 \times 10^{-3}$) was calculated using the total 2027 bp available.

Phylogenetic reconstruction: A *cytB* nj tree constructed with wild and domestic reference samples ($n = 202$, Figure 1a) revealed Turkish and Israeli sheep separated into five distinct lineages. The majority of animals ($n = 191$) grouped into the previously described clades A, B, (Hiendleder *et al.* 1998), and C (Pedrosa *et al.* 2005). Two animals from the Turkish Morkaraman (MK) breed shared a haplotype, which branched away from all other domesticates and was of the same lineage as *O. vignei*, *O. orientalis*, and *O. ammon*. Similarly divergent haplotypes have recently been reported in a single Karachai animal (Tapio *et al.* 2006) and a Mongolian individual (Guo *et al.* 2005). To facilitate direct sequence comparison, a 422-bp region of the highly variable control region (hv-mtCR) was sequenced and trimmed (DQ852280–DQ852289) as described by Tapio *et al.* (2006). Analysis revealed that the MK sequence (this study) displayed only one and two nucleotide differences when compared with the Karachai and Mongolian sequences, respectively ($K =$ 0.2 and 0.5%). This level of diversity is equal to or less than the amount observed within clades when using this highly variable region of the mtDNA molecule (Tapio *et al.* 2006; $K = 0.4$–0.9%). The two Morkaraman sequences were therefore termed lineage D. Figure 1a revealed a fifth distinct grouping. Four animals (AW, $n = 3$; TJ, $n = 1$, Table 1) formed a separate cluster that linked clades A and C. These assembled with a published sequence, which previously appeared as a divergent branch of clade C (Karayaka from Pedrosa *et al.* 2005). This newly identified lineage was termed E and was supported by a bootstrap value of 98%. A second nj tree that considered mtCR-*cytB* data in the absence of a wild outgroup sequence presented a similarly shaped result with higher bootstrap values (bootstrap values for this tree are shown in parentheses on the *cytB* tree, Figure 1a). To further test tree topology, the mtCR-*cytB* data were used for construction of a Bayesian-derived consensus tree. The same tree topology was observed where each of the four lineages containing multiple members (Table 1) formed distinct clusters with high support (A, 1.00; B, 0.70; C, 1.00; E, 1.00). The fifth lineage (D) originated from within lineage A rather than between A and C as seen in Figure 1a.

To evaluate the relationship within and between haplogroups, the *cytB*-mtCR data set was used in the construction of a median joining network (Figure 1b). Five clearly defined haplogroups were observed and designated HA, HB, HC, HD, and HE. HA, HB, and HC reflect the previously defined nomenclature for sequences belonging to clades A, B, and C while HD and HE were named in reference to clusters D and E (Figure 1a). The three major haplogroups HA, HB, and HC form general star-burst clusters where smaller derivative groups branch from a central haplotype, consistent with population expansions. HB has two large central haplotypes with 24 and 17 members, comparable to the two central haplotypes that were identified in clades B and A by Pereira *et al.* (2006) when they collated the available published mt control region data.

The distribution of haplogroups across sheep breeds is summarized in Table 1. The Sakiz (limited to HB) and Karya breeds (HA and HB) have restricted memberships, while the remaining breeds contain three (for KK, NZ, and CC) or four (for MK, TJ, and AW) haplogroups. For eight of the nine breeds sampled, the most prevalent haplogroup was HB. Only Norduz (NZ) differed in this regard, with the majority of animals having HA sequences.

The average number of nucleotide differences (D) and sequence diversity (K) between previously described clades (A, B, and C) was compared with those defining the new haplogroups D and E to determine if they represent newly identified lineages. Haplogroups HA and HB were separated by an average of 11.72 substitutions, similar to the value reported by Meadows *et al.* (2005) for the same clusters ($D = 11.41$). For HA–HD ($D = 15.93$) and HA–HE ($D = 13.89$), the average number of observed nucleotide differences was higher, while for HC–HE the value was slightly lower ($D = 9.80$). Sequence diversity, in terms of K, for the same lineage comparisons showed the same trend ranging from a low of 0.48% for HC–HE to the highest variation of 0.79% for HA–HD. Inspection of mismatch distributions was

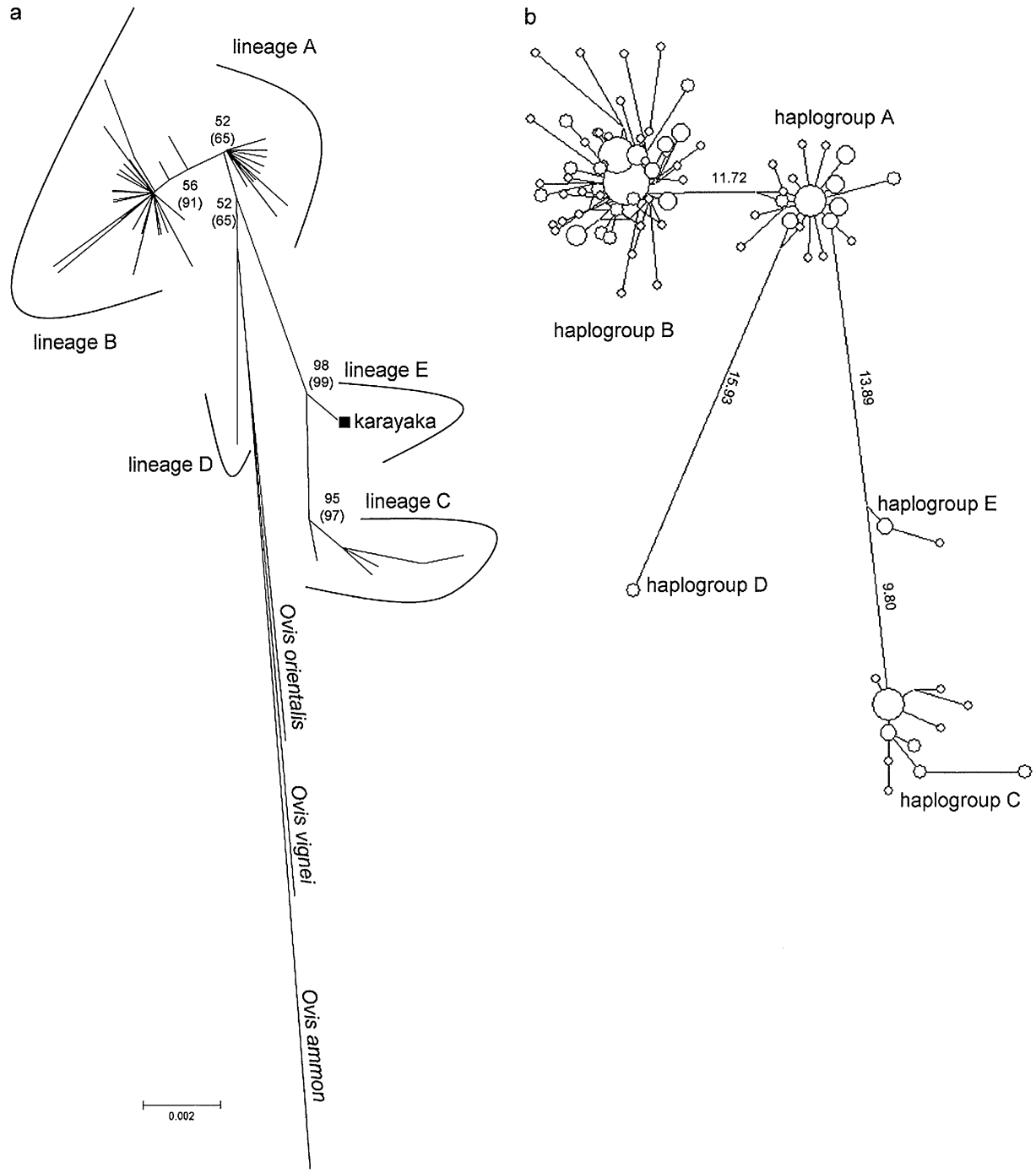

FIGURE 1.—Five mtDNA lineages illustrated with two types of phylogenetic tree. (a) Neighbor-joining tree showing *O. aries* lineages in relation to wild sheep using *cytB* sequence (967 bp). Analysis of animals from nine domestic breeds ($n = 197$) was supplemented with wild Ovis species and a divergent haplotype previously identified from the Karayaka breed (PEDROSA *et al.* 2005). Additional data on the relationship between wild Ovis species are presented elsewhere (HIENDLEDER *et al.* 2002; BUNCH *et al.* 2006). Bootstrap values are indicated on cluster nodes; values in parentheses were taken from a similar tree constructed using the mtCR-*cytB* data set (2027 bp). (b) Weighted median-joining network showing mt haplotypes. Node size is proportional to haplotype frequency and the mutational differences between haplogroups are proportional to branch length (*D* shown on branch). The smallest node is representative of one animal.

also used to determine if haplogroups D and E were sufficiently divergent to represent separate domestic clades (Figure 2). Analysis of the total data set (mtCR-*cytB*; $n = 197$) revealed a tetramodal curve (Figure 2a). The three haplogroups with more than two haplotypes were plotted individually (Figure 2, b–d) to search for signals of clade substructure. Smooth bell-shaped distributions were observed for HA and HB, indicating a lack of substructure. Haplotypes from adjacent haplogroups were then analyzed together to explore the

TABLE 1

Summary of sheep breed and mitochondrial haplogroup phylogenetic variation

	Haplogroup					Total
	HA	HB	HC	HD	HE	(*n*)
Sequences	45	119	27	2	4	197
Variable sites[a]	27	72	16	0	2	116
Haplotypes	19	52	11	1	2	85
Nucleotide diversity[b] (π)	1.06	1.64	1.15	0.00	0.49	
Haplogroup membership (%)						
Karakas (KK)	20.0	60.0	20.0			20
Morkaraman (MK)	26.3	57.9	5.3	10.5		19
Tuj (TJ)	18.8	43.8	31.2		6.2	16
Karya (KR)	12.5	87.5				24
Norduz (NZ)	46.8	26.6	26.6			15
Cine Capari (CC)	14.3	50.0	35.7			14
Karayaka (KY)	40.0	60.0				15
Sakiz (SZ)		100.0				17
Awassi (AW)	26.3	54.4	14.0		5.3	57

The geographic origin and population structure of each breed are described in MATERIALS AND METHODS.

[a] Variable sites are a combination of indel and single nucleotide polymorphism.

[b] Nucleotide diversity ($\times 10^{-3}$) was calculated excluding indels.

resultant shift in mismatch distribution. Combined analysis of HA and HD showed pairwise combinations with >15 mismatches (Figure 2e). This was more than the observed separation between HA from HB (Figure 2e), confirming that HD is clearly distinct from other groups. A similar combined analysis was performed using HC and HE (Figure 2f). This resulted in pairwise combinations with nine or greater mismatches. This was not observed when HC sequences were analyzed in isolation (Figure 2d), supporting HE as a separate group to HC. Finally, *cytB* genetic diversity (*K*) was calculated between each domestic haplogroup and the closest wild Ovis species as indicated by the nj tree (Figure 1a). The most divergent sequence to *O. orientalis* was HC ($K = 2.34\%$), while the most similar was HA ($K = 1.60\%$). All *K*-values are presented in Table 2.

Population expansion: The star-burst patterns observed for some haplogroups (Figure 1b) and differences in the shape of each mismatch distribution (Figure 2) prompted an analysis of population demography. Calculation of Fu's F_s statistic was performed for haplogroups A, B, and C as a diagnostic of past population growth (EXCOFFIER *et al.* 2005). The presence of population expansions was supported for HA ($F_s = -13.2$) and HB ($F_s = -26.2$) with high significance ($P < 0.02$) but was rejected for HC (HC, $F_s = -2.8$, $P = 0.06$). This appears to match the observed smooth bell-shaped mismatch curves seen for HA and HB and the ragged bimodal distribution observed for HC (Figure 2). The irregular curve of HC is consistent with previous findings (PEDROSA *et al.* 2005). The mismatch distributions for HA and HB were tested against the sudden expansion model and found not to differ from the fitted model (HA, $P = 0.14$; HB, $P = 0.90$). This allowed estimation of the time in years since the commencement of expansion (t) and an associated 95% confidence interval. For HA ($\tau = 1.06$), $t = 21{,}000$ (12,800–33,000) years while for HB ($\tau = 1.14$), $t = 23{,}500$ (17,300–32,300) years.

DISCUSSION

In the search for new lineages of domestic livestock, it makes sense to search where the gene pool is expected to be the most diverse, that is, the center of animal domestication. The rationale follows that, as human populations radiate to colonize new sites, they would take with them only a subset of domestic livestock, leading to the expectation that, with increased geographic distance, there would follow a decrease or loss of genetic lines. For sheep, this domestication center has been located to the Levant, which has been archaeologically dated to the Neolithic agricultural revolution. This study reports increased levels of sheep nucleotide diversity in this region, lending support to its geographic role in domestication and indicating a more complicated pattern of ovine domestication than previously assumed.

The evolutionary history of 197 sequences from eight Turkish and one Israeli breed were examined using a range of phylogenetic methods. Neighbor joining and Bayesian trees along with a phylogenetic network diagram revealed five distinct groups, named HA, HB, HC, HD, and HE. Haplogroups HD and HE clustered independently of sequences that corresponded to the previously described clades A, B, and C, suggesting the identification of previously unidentified and distinct maternal clades. In the absence of an agreed upon set of molecular parameters to define a clade, diversity indexes defining clusters HD and HE were compared to an established clade pair, A and B. The average number of mutational changes between groups (Figure 1b) showed that a similar number of mutational positions separated HE from HC ($D = 9.8$) as were observed between the accepted clades HA and HB ($D = 11.7$). A higher number of observed mutations ($D = 15.9$) was observed between HA and HD. These estimates should be considered preliminary and would be expected to contract as more animals carrying these rare haplotypes (HD and HE) are identified. A similar trend was observed in sequence diversity (K) values (Table 2). Taken in conjunction, these comparisons support the conclusion that haplogroups HD and HE are sufficiently divergent to be considered new ovine maternal clades, termed D and E.

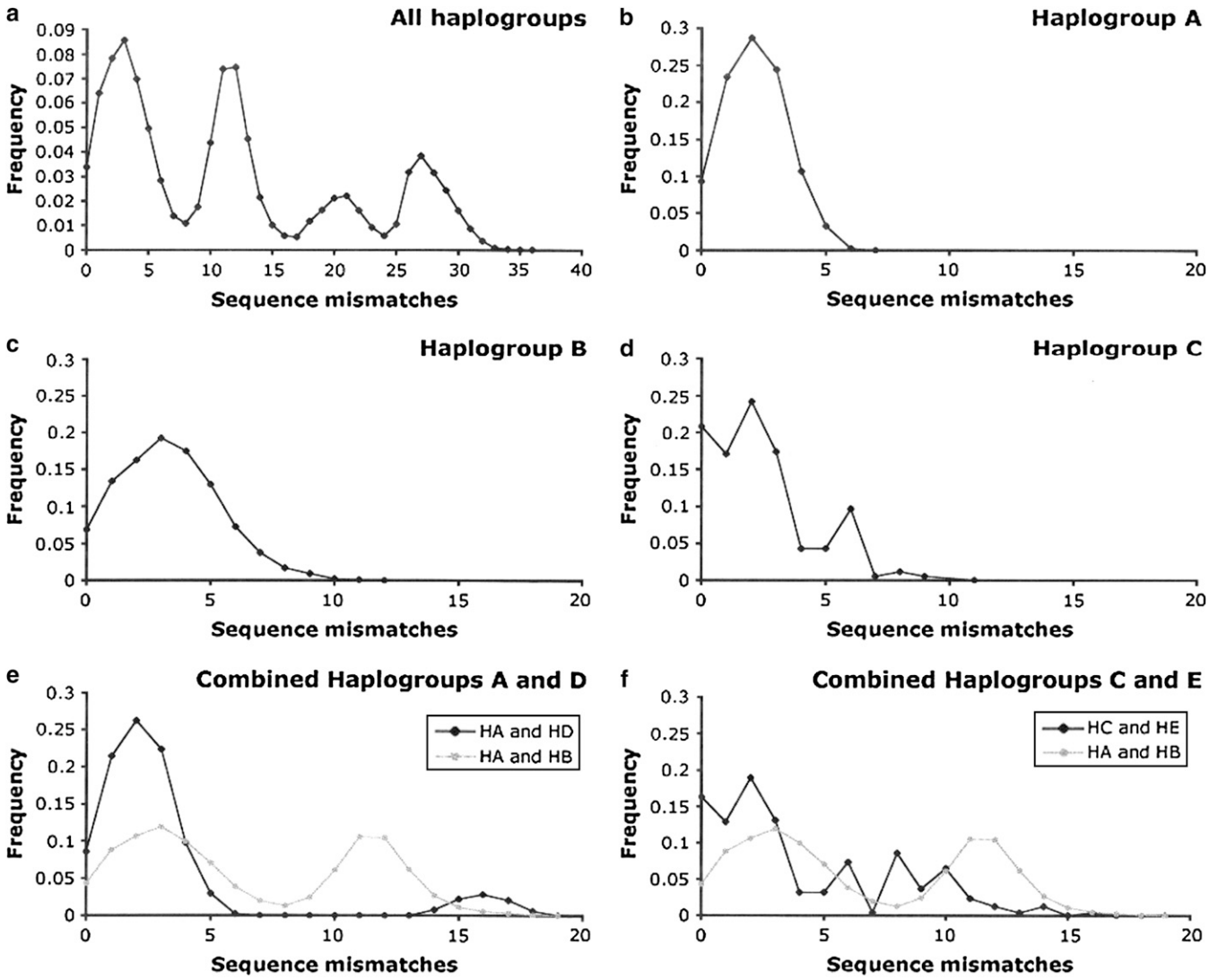

FIGURE 2.—Mismatch distributions within and between clades using the mtCR-*cytB* data set (2027 bp). Analysis of all 85 haplotypes revealed a tetramodal distribution, which is clear evidence for multiple ovine lineages (a). Haplogroups A, B, and C were analyzed separately (b, c, and d, respectively). Haplogroups were mixed to investigate the relationship between adjacent sequence sets. (e) The distribution following mixing of HA and HD. (f) The distribution following mixing of HC and HE.

Others have noted the rare possibility that divergent low-frequency mtDNA haplotypes such as HD and HE may in fact be nuclear mitochondrial pseudogenes (numts) (PARR *et al.* 2006). To confirm that the sequence variation observed here was not due to the numts, two animals from each haplogroup were amplified and sequenced using separate primer sets for hvmtCR ($n = 10$). All hvmtCR sequences lacked

TABLE 2

Genetic diversity among five clades of domestic sheep and their closest wild relative

	Clade A	Clade B	Clade C	Clade D	Clade E	*O. orientalis*
Clade A						
Clade B	0.58 (0.31)					
Clade C	0.98 (1.08)	1.37 (1.29)				
Clade D	0.79 (0.77)	1.07 (0.98)	1.28 (1.55)			
Clade E	0.66 (0.70)	1.05 (0.91)	0.49 (0.44)	1.06 (1.16)		
O. orientalis	(1.60)	(1.80)	(2.34)	(1.86)	(1.99)	

Nucleotide substitutions per site (K%) were calculated in the absence of indels using either the combined mtCR and *cytB* sequence (2027 bp) or the *cytB* sequence alone (967 bp, reported in parentheses).

heteroplasmy and so it is extremely unlikely that the divergent haplogroups HD and HE are numts.

Two recent studies designated clade membership using either sequence from the control region (Guo *et al.* 2005) or this and independent analysis of cytochrome *B* (Pedrosa *et al.* 2005). Both surveys identified significant sequence divergence within clades. Inspection of Turkish breeds revealed two haplotypes branched apart from other clade C sequences with high support (Pedrosa *et al.* 2005). In a third study, analysis of Chinese breeds showed two divergent clade A groups (termed subclade a1 and a5), which were separated by approximately the same number of substitutions as clades A and B (Chen *et al.* 2006). Utilization of cytochrome *B* sequence from this and previous work (Pedrosa *et al.* 2005) allowed joint analysis, which revealed the divergent Turkish haplotype observed in Karayaka (Pedrosa *et al* 2005) clearly clustered with clade E (Figure 1a). Joint analysis of subclade a1 and a5 was performed using the overlapping mtDNA control region sequence common to both studies (446 bp, this study and Chen *et al.* 2006). Alignment revealed subclade a1 grouped with lineage A sequences; however, the subclade a5 sequence grouped separately to all other lineages. Significantly, however, alignment of available cytochrome *B* sequence from a5 (DQ309021) revealed that it grouped with the clade A sequences reported in this study (average $K = 0.003$ substitutions/site). These analyses showed that subclade a5 was not equivalent to clade D or E, but highlighted the need to assay multiple regions of the mtDNA molecule. This study employed a total of 2027 bp of sequence sampled from the control region, tRNA-Phe, 12s rRNA gene, and cytochrome *B* gene. Interpretations therefore arise from inspection of segments of the molecule that are under four forms of evolutionary constraint and therefore display different rates of mutational change (Pesole *et al.* 1999). The result was the identification of five clearly defined haplogroups (Figure 1b), which may not have been possible using inspection of a single region alone.

The frequency of haplogroups was determined within each of the nine breeds under investigation (Table 1). The majority of breeds (six of nine) demonstrated high levels of clade diversity, with three or more lineages represented. An independent sampling of four of the same breeds (TJ and KY and two forms of Akkaraman KK and NZ) was recently reported (Pedrosa *et al.* 2005), allowing for a comparison of haplogroup frequencies. Interestingly, the observed frequencies varied considerably. In the case of Karayaka, Pedrosa *et al.* (2005) observed 35% breed membership to clade C, whereas this clade was not identified in our samples (Table 1). Similarly, the Tuj animals in this study were sampled from a single flock, with familiar relationships unknown, and yet showed a higher clade membership to clade C (31% in Table 1, 7% in Pedrosa *et al.* 2005) and a presence in clade E (6.2%). These within-breed variations likely reflect the small sample sizes assayed. In each study, these four breeds are defined by between 15 and 20 animals. However, these discrepancies also indicate the highly heterogeneous nature of these breeds and indicate that caution should be used when interpreting interbreed relationships.

Clades A and B were evidenced to have undergone population expansions. Dating these events is important as it infers a time window during which the population of each clade became significantly more numerous, one of the telltale signs of domestication. For clades A and B, the most recent date for expansion was placed at 12,800 and 17,300 years ago, respectively. While the date given for clade A is close to the figure calculated for domestication on the basis of the fossil record (9000 years; Ryder 1984), the estimate for clade B predates the shift of humans from hunter-gathers to subsistence farmers, indicating that a large and diverse pool of predomesticated animals may have been available. Other studies have not directly tested for this possibility as the timing of the first and oldest expansion was assumed to equal that of domestication. As a result, clade expansion times in both sheep (Tapio *et al.* 2006) and goat (Luikart *et al.* 2001) were dated as more recent than 10,000 years before present. The possibility remains that population expansions predate domestication; however, caution is required before firm conclusions can be drawn. The calibration point used in this study to derive an Ovis-specific mutation rate is surrounded by significant error (Bunch *et al.* 2006). In addition, a single and constant mutation rate was assumed, which may lead to an overestimate in molecular dating (Ho and Larson 2006). The resulting expansion estimates (12,800–17,300 years) are much smaller than recent findings for another domestic animal (pig), which showed evidence of demographic expansions hundreds of thousands of years before present (Fang and Andersson 2006). In keeping with past sheep estimates (Chen *et al.* 2006; Tapio *et al.* 2006), the dates derived in this work provide support for clades A and B being the result of separate domestication events, as it is unlikely that such genetic diversity would have been maintained in populations where introgression was possible. It is anticipated that, with a greater sampling of animals from this region, more clade C, D, and E sequences would become available and that similar calculations could then be performed with these rarer sequence sets.

The identification of additional maternal lineages offered the opportunity to revisit the relationship between domestic sheep and various wild species. Ovis species are split into four phylogenetic groups, which span Eurasia (Moufloniform, Argaliform I and II) and America (Pachceriform) (Bunch *et al.* 2006). The Moufloniforms include domestic sheep and three wild species, *O. vignei*, *O. orientalis*, and *O. musimon*. Of these wild sheep, the European moufon (*O. musimon*) is

considered a feral domesticate of the Eastern mouflon (*O. orientalis*) that was transported into Europe from Asia after the Neolithic period. Ovis species in overlapping geographic ranges are known to hybridize and produce fertile offspring (reviewed in Hiendleder *et al.* 2002) and, in particular, *O. orientalis* has been proposed as the progenitor to Asian sheep clades (Hiendleder *et al.* 2002). A phylogenetic tree drawn using 967 bp of cytochrome *B* sequence (Figure 1a) allowed for a direct comparison between the wild and newly identified domestic lineages. While clade D was observed to branch with wild animals, it did not form a cluster with the eastern mouflon, urial (*O. vignei*), or argali (*O. ammon*) sheep. The positioning of clade D should be treated with caution as it showed the equal lowest support in the nj *cytB* tree (Figure 1a) and branched separately to wild sequence when a portion of the mtDNA control region was analyzed (Tapio *et al.* 2006). In addition, both network analysis (Figure 1b) and Bayesian inference indicated that the origin of the single lineage D sequence was from within lineage A, not from a branch located between lineage A and E as in Figure 1a. Inspection of the level of nucleotide diversity separating clade D and *O. orientalis* ($K = 1.86\%$) was larger than the value observed between the most divergent combination of domestic haplogroups HC–HD ($K = 1.55\%$, Table 2). This leaves the wild link to this and the other clades (A, C, and E) unresolved.

The observed maternal origins of sheep are beginning to match the emerging pattern observed in other domestic species. As more animals are examined, especially from regions archaeologically evidenced as centers for domestication, signatures of additional domestication events are being identified. For sheep, this has increased from two lineages first described in 1996 (Wood and Phua *et al.* 1996) and 1998 (Hiendleder *et al.* 1998) to the five now identified in the Near East. Interestingly, even with the discovery of two new clades, no extant wild Ovis progenitor has been identified. The European Mouflon (*O. musimon*) carries HB (Hiendleder *et al.* 2002; Meadows *et al.* 2005); however, it is a neolithic feral domesticate. It is possible, however, that domestications sampled diverse Mouflon-iformic populations, which have subsequently ceased to exist. Perhaps the next phase in mtDNA discovery should be aimed at ancient DNA from the Fertile Crescent to determine if the wild sheep of today do in fact have an ancient link to modern domesticates or, if like other species, their progenitor is now extinct.

The authors thank the following people for their contribution of blood samples: Murat Yilmaz, Onur Yilmaz, and Seval Ozdemir (Adnan Menderes University, Turkey) for SZ, CC, and KR; Turgut Aygun and Ayhan Yilmaz (Yuzuncu Yil University, Turkey) for NZ and KK; Hakki Emsen and Ebru Emsen (Ataturk University, Turkey) for MK and TJ; and Zafer Ulutas and Emre Sirin (Gaziosmanpasa University, Turkey) for KY samples. Pelin Binbas (Adnan Menderes University, Turkey) is thanked for Turkish DNA extractions and the team of the Ein Harod Awassi flock, Israel, for animal handling. The authors also thank the two anonymous reviewers. Bioinformatic support was provided by Wes Barris (Commonwealth Scientific and Industrial Research Organization, Australia). J.R.S.M was supported by a Ph.D. scholarship from Meat & Livestock Australia.

LITERATURE CITED

Bandelt, H-J., P. Forster and A. Rohl, 1999 Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16:** 37–48.

Bunch, T. D., C. Wu, Y.-P. Zhang and S. Wang, 2006 Phylogenetic analysis of Snow Sheep (*Ovis nivicola*) and closely related taxa. J. Hered. **97:** 21–30.

Chen, S-Y., Z-Y. Duan, T. Sha, J. Xiangyu, S-F. Wu *et al.*, 2006 Origin, genetic diversity, and population structure of Chinese domestic sheep. Gene **376:** 216–223.

Excoffier, L., G. Laval and S. Schneider, 2005 Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol. Bioinform. Online **1:** 47–50.

Fang, M., and L. Andersson, 2006 Mitochondrial diversity in European and Chinese pigs is consistent with population expansions which occurred prior to domestication. Proc. Biol. Sci. **273:** 1803–1810.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Guo, J., L. X. Du, Y. H. Ma, W. J. Guan, H. B. Li *et al.*, 2005 A novel maternal lineage revealed in sheep (*Ovis aries*). Anim. Genet. **36:** 331–336.

Hasegawa, M., H. Kishino and T. Yano, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22:** 160–174.

Hiendleder, S., K. Mainz, Y. Plante and H. Lewalski, 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. J. Hered. **89:** 113–120.

Hiendleder, S., B. Kaupe, R. Wassmuth and A. Janke, 2002 Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proc. Biol. Sci. **269:** 893–904.

Ho, S. Y., and G. Larson, 2006 Molecular clocks: when times are a-changin'. Trends Genet. **22:** 79–83.

Huelsenbeck, J. P., and F. Ronquist, 2001 MRBAYES: Bayesian inference of phylogeny. Bioinformatics **17:** 754–755.

Joshi, M. B., P. K. Rout, A. K. Mandal, C. Tyler-Smith, L. Singh *et al.*, 2004 Phylogeography and origin of Indian domestic goats. Mol. Biol. Evol. **21:** 454–462.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:** 150–163.

Larson, G., K. Dobney, U. Albarella, M. Fang, E. Matisoo-Smith *et al.*, 2005 Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science **307:** 1618–1621.

Legge, T., 1996 The beginning of caprine domestication, pp. 238–262 in *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, edited by D. R. Harris. Smithsonian Institution Press, New York.

Li, W-H., 1997 *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

Luikart, G., L. Gielly, L. Excoffier, J. D. Vigne, J. Bouvet *et al.*, 2001 Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc. Natl. Acad. Sci. USA **98:** 5927–5932.

Meadows, J. R. S., K. Li, J. Kantanen, M. Tapio, W. Sipos *et al.*, 2005 Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. J. Hered. **96:** 494–501.

Miller, S. A., D. D. Dykes and H. F. Polesky, 1988 A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. **16:** 1215.

Parr, R. L., J. Maki, B. Reguly, G. D. Dakubo, A. Aguirre *et al.*, 2006 The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. BMC Genomics **7:** 185.

Pedrosa, S., M. Uzun, J. J. Arranz, B. Gutierrezz-Gil, F. San Primitivo *et al.*, 2005 Evidence of three maternal lineages in

near eastern sheep supporting multiple domestication events. Proc. Biol. Sci. **272:** 2211–2217.

Pereira, F., S. J. M. Davis, L. Pereira, B. McEvoy, D. G. Bradley *et al.*, 2006 Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. Mol. Biol. Evol. **23:** 1420–1426.

Pesole, G., C. Gissi, A. De Chirico and C. Saccone, 1999 Nucleotide substitution rate of mammalian mitochondrial genomes. J. Mol. Evol. **48:** 427–434.

Posada, D., and K. A. Crandall, 1998 Modeltest: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Rogers, A., 1995 Genetic evidence for a Pleistocene population explosion. Evolution **49:** 608–615.

Rogers, A. R., and H. Harpending, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9:** 552–569.

Ronquist, F., and J. P. Huelsenbeck, 2003 MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572–1574.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP: DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Ryder, M. L., 1984 Sheep, pp. 63–84 in *Evolution of Domesticated Animals*, edited by I. L. Mason. Longman Group, London/New York.

Sardina, M. T., M. Ballester, J. Marmi, R. Finocchiaro, J. B. van Kaam *et al.*, 2006 Phylogenetic analysis of Sicilian goats reveals a new mtDNA lineage. Anim. Genet. **37:** 376–378.

Scherf, B. D. (Editor), 2000 *World Watch List for Domestic Animal Diversity*, Ed. 3. Food and Agriculture Organization of the United Nations, Rome.

Schneider, S., and L. Excoffier, 1999 Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics **152:** 1079–1089.

Swofford, D. L., 2003 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tamura, K., and M. Nei, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512–526.

Tapio, M., N. Marzanov, M. Ozerov, M. Cinkulov, G. Gonzarenko *et al.*, 2006 Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. Mol. Biol. Evol. **23:** 1776–1783.

Wood, N. J., and S. H. Phua, 1996 Variation in the control region sequence of the sheep mitochondrial genome. Anim. Genet. **27:** 25–33.

Communicating editor: C. Haley