ORIGINAL ARTICLE

Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel

JRS Meadows[1,3], S Hiendleder[2] and JW Kijas[1]

[1]CSIRO Livestock Industries, St Lucia, Queensland, Australia and [2]School of Agriculture, Food and Wine & Research Centre for Reproductive Health, School of Animal and Veterinary Science, University of Adelaide, Roseworthy, South Australia, Australia

Five haplogroups have been identified in domestic sheep through global surveys of mitochondrial (mt) sequence variation, however these group classifications are often based on small fragments of the complete mtDNA sequence; partial control region or the *cytochrome B* gene. This study presents the complete mitogenome from representatives of each haplogroup identified in domestic sheep, plus a sample of their wild relatives. Comparison of the sequence successfully resolved the relationships between each haplogroup and provided insight into the relationship with wild sheep. The five haplogroups were characterised as branching independently, a radiation that shared a common ancestor 920 000 ± 190 000 years ago based on protein coding sequence. The utility of various mtDNA components to inform the true relationship between sheep was also examined with Bayesian, maximum likelihood and partitioned Bremmer support analyses. The control region was found to be the mtDNA component, which contributed the highest amount of support to the tree generated using the complete data set. This study provides the nucleus of a mtDNA mitogenome panel, which can be used to assess additional mitogenomes and serve as a reference set to evaluate small fragments of the mtDNA.
Heredity (2011) **106,** 700–706; doi:10.1038/hdy.2010.122; published online 13 October 2010

Introduction

Domesticated sheep have had an important part in human settlement, providing a farmed source of food, wool and hide since the Neolithic Agricultural revolution approximately 8000–9000 years ago (Ryder, 1984). From these early beginnings, more than 1400 breeds of sheep are currently recognised (Scherf, 2000). Elucidating the origins of domestic sheep and their human-mediated global migrations has, in large part, been examined through analysis of mitochondrial (mt) DNA. Mitochondrial DNA are useful in answering phylogenetic questions as they follow a maternal mode of inheritance, have a high copy number and have a greater rate of substitution on average than nuclear genes, making them particularly useful for resolving intra-species branching (Moore, 1995). Mitochondrial DNA variation studies have revealed five haplogroups (termed HA, HB, HC, HD and HE) in domestic sheep sampled from a range of geographically dispersed locations. Haplogroups HD and HE were identified most recently and are also the rarest, to date found only in sheep from the Caucasus and Turkey (Tapio et al., 2006; Meadows et al., 2007). Haplogroup HC has the next most restricted distribution with examples located in Asia, the Fertile Cresent, Caucasus and the Iberian Peninsula (Guo et al., 2005;

Pedrosa et al., 2005; Pereira et al., 2006; Tapio et al., 2006; Meadows et al., 2007). HA and HB are the most frequently identified haplogroups. These variants were first documented by Wood and Phua (1996) and classified by Hiendleder et al. (1998), but have since been located in every geographic region where *Ovis aries* have been sampled.

The five haplogroups have been defined based on limited segments of the mtDNA, usually a control region fragment or the *cytochrome B* gene. These two fragments represent less than 12% of the complete genome, meaning care is required when drawing conclusions concerning divergence times and relatedness, given the mutations accumulate at different rates between the various components of the mitogenome (reviewed by Bandelt et al., 2006). Analysis based on short regions of the mtDNA may also generate trees, which lack statistical robustness in their branching topology. A clear example is the position of haplogroup HD when compared between the studies (Tapio et al., 2006; Meadows et al., 2007). The relative position of this haplogroup differs and in each case the statistical support associated with its position is poor.

This study aimed to sequence the complete mt genome of representatives from each of the five haplogroups identified in domestic sheep, plus a subset of wild sheep, and in doing so, form the nucleus of a global ovine mtDNA reference panel. This panel was then used to resolve the phylogenetic relationships between wild and domestic sheep for the first time using a whole genome mtDNA tree. Access to the full genome enabled the comparison and evaluation of constituent mtDNA coding and non-coding fragments for tree construction

Correspondence: Dr JW Kijas, CSIRO Livestock Industries, 306 Carmody Road, St Lucia 4067, Australia.
E-mail: James.Kijas@csiro.au
[3]Current address: Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 75123, Sweden.

and phylogenetic resolution. Finally, the evolutionary divergence between the five recognised haplogroups was revisited to examine whether all represent independent lineages.

Materials and methods

Sample selection
Five haplogroups HA, HB, HC, HD and HE have previously been reported in domestic sheep based on a concatemer of partial control region sequence, tRNA-Phe, 12s rRNA (1060 bp) and cytochrome B sequence (967 bp) (Meadows et al., 2007). To select two animals, representative of each haplogroup, all samples previously sequenced by the authors (Meadows et al., 2005, 2007) were pooled ($n = 318$) and a conservative, weighted median-joining diagram constructed using Network 4.1.1.2 (http://www.fluxus-engineering.com) (Supplementary Figure 1). Network construction was as described previously (Meadows et al., 2007). The animals selected for mitogenome sequence analysis are described in Table 1. For haplogroups HA, HB and HC, individuals were selected, which carried the central and most frequently observed haplotype (Supplementary Figure 1). Animals cl122 and r359 were selected to represent HA, kk1 and kk2 for HB and mk4 and kk12 for HC. HD was composed of a single haplotype containing two samples (mk3 and mk9), both of which were taken forward into sequencing. HE was represented by aw25 and tj6. A further six ovine samples were included in the study, a single O. ammon (h77), three O. vignei (h75, h76 and h78) and two O. musimon (h1, h2). Although O. musimon are considered examples of feral domesticates, they were termed 'wild' and grouped with the other Ovis spp. data set.

Mitochondrial DNA sequencing
Supplementary Table 1 describes the primer pairs designed using ovine reference sequence AF010406 (Hiendleder et al., 1998), which were employed to amplify the 18 fragments of the mt. PCR products from the 10 domestic and 6 wild animals were produced using the conditions of Meadows et al. (2007) and were sequenced directly in both the forward and reverse direction using Big Dye Terminator v3.1 chemistry (Applied Biosystems, Foster City, CA, USA). Sequence reads were collected with a 3130xl Genetic Analyser (Applied Biosystems) and aligned with Sequencher 4.2.2 (Gene Codes, Ann Arbor, MI, USA). A nested sequencing reaction was performed on each DNA sample to test for nuclear inserts of DNA using primer pairs and methods of (Supplementary Table 1; Tapio et al., 2006). Sequencing, visualisation and alignment of this fragment was performed as above.

Sequence comparison and phylogenetic inference
The repeat unit located within the mt control region was removed before phylogenetic inference because of its known heteroplasmic behaviour (Hiendleder et al., 2002). The aligned sequence of each complete mtDNA genome was fractionated into 17 components; 13 protein-coding genes, 2 srRNAs, control region and a concatemer of transfer RNAs. In addition, data sets were generated, which contained all of the protein coding genes (pr-mtDNA), the two srRNAs (s-mtDNA) or the full sequence of the mtDNA genome (gen-mtDNA). Sequence variation, calculated as nucleotide diversity (π), nucleotide differences (D) and number of substitutions per site (K), were calculated in DnaSP 4.20.1 (Rozas et al., 2003). MODELTEST 3.0.6.6 (Posada and Crandall, 1998) implemented in PAUP* 4b10-ppc-macosx (Swofford, 2003) used hierarchical likelihood ratio tests to identify the specific nucleotide substitution model for each of the individual and concatenated data sets. The HKY ($ND3$, $ND6L$, COI, $COII$, $ATPase6$, $ATPase8$, t-mtDNA and s-mtDNA) or HKY $+ \Gamma$ ($ND1$, $ND2$, $ND4$, $ND4L$, $ND5$, $COIII$, Cyt b and pr-mtDNA) models were identified to best fit most data sets, whereas the gen-mtDNA concatemer was best described by the TrN $+ I + \Gamma$ model. The appropriate substitution model was applied to each

Table 1 Mitogenome sequence derived from 10 domestic and 6 wild sheep samples

Group membership	Sample name	Common name	Sample location	Length (bp)	Accession number
O. aries[a]					
Haplogroup A	cl122	Merino	Australia	16 617	HM236174
	R359	Romney	Australia	16 617	HM236175
Haplogroup B	kk1	Karakas	Turkey	16 616	HM236176
	kk2	Karakas	Turkey	16 616	HM236177
Haplogroup C	kk12	Karakas	Turkey	16 620	HM236178
	mk4	Morkaraman	Turkey	16 620	HM236179
Haplogroup D	mk3	Morkaraman	Turkey	16 616	HM236180
	mk9	Morkaraman	Turkey	16 616	HM236181
Haplogroup E	aw25	Awassi	Israel	16 620	HM236182
	tj6	Tuj	Turkey	16 620	HM236183
Other Ovis spp.[b]					
O. musimon	h1	Mouflon	Germany	16 616	HM236184
	h2	Mouflon	Germany	16 616	HM236185
O. ammon	h77	Argali	Kazakhstan	16 613	HM236188
O. vignei	h75	Urial	Kazakhstan	16 694	HM236186
	h76	Urial	Kazakhstan	16 694	HM236187
	h78	Urial	Kazakhstan[c]	16 696	HM236189

[a]Domestic sheep DNA was extracted as described in Meadows et al. (2005).
[b]Other Ovis spp. DNA was extracted as per Hiendleder et al. (2002).
[c]Sample h78 was from Turkmenistan, sampled in Kazakhstan.

Mitogenome sequences of both wild and domestic sheep
JRS Meadows et al
702

data set in PAUP* to generate bootstrap (1000 replications)-supported maximum likelihood trees. All data sets were also examined in a Bayesian framework with MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). The general time-reversible nucleotide substitution model was used, and rate variation across sites was modelled using a gamma distribution approximated using four rate categories. The analysis assumed that a proportion of sites are likely to be invariable. This was modelled assuming a uniform distribution between the extremes in which all sites are invariable and no sites are invariable. The analysis itself was run for 1 million iterations that ensured the s.d. of split frequencies dropped below 0.01 and the Markov chain Monte Carlo algorithm reached convergence. TreeRot v3 (Sorenson and Franzosa, 2007) was used to perform a partitioned Bremer support (PBS) analysis. This aims to measure the support that various sequence subsets contribute to the tree generated using the full data set. The 17 components of the mtDNA genome described above were treated as separate partitions, and PBS analysis performed on each either with, or without, inclusion of the wild sheep sequence.

Time to most recent common ancestor

Database *cytB* sequence from the extinct *Myotragus balearicus* (AY380560) was obtained to facilitate the molecular dating of *Ovis* spp. using a fossil calibration reference. *Myotragus* was a goat-like member of the *Caprinae* subfamily, which were isolated on the Balearic Islands following the opening of the Strait of Gibraltar 5.35 Myr ago (reviewed in Lalueza-Fox *et al.*, 2005) allowing this date to be used as a fixed age in a *Bovidae cytB* tree. The node joining *O. ammon* (h77) to the remaining ovine samples was dated (2.13 ± 0.29 Myr) and used as a calibration point (CP) in a tree constructed from the 12H strand protein-coding genes. Tajima's relative rate test (Tajima, 1993) showed the molecular clock was violated in both sequence sets (*cytB* and 12H strand protein concatemer), leading to the implementation of nonparametric rate smoothing (Sanderson, 1997) in r8s v1.7 (Sanderson, 2003) for divergence time calculation. *Bos taurus* (NC_006853) sequence was used to root each tree and pruned before the calculation of the date. Confidence intervals were determined following the r8s bootstrap kit scripts of Torsten Eriksson (www.bergianska.se/index_forskning_soft.html).

Results

Mitogenome sequence diversity

A total of 16 complete mtDNA genomes were sequenced and each represented a unique haplotype. The mitogenomes from domestic sheep ranged in size from 16 616 to 16 620 bp and in wild animals from 16 613 to 16 696 bp. The variation in size was largely due to a 75 or 76 bp repeat motif, which is located within the control region (Hiendleder *et al.*, 2002). The majority of domestic sheep contained four copies of a 75 bp repeat unit, resulting in a mitogenome of 16 616 or 16 617 bp (Table 1). The mtDNA genome in four domestic sheep was slightly longer (16 620 bp), due to the presence of one 75 bp repeat followed by three 76 bp repeats. Both the mouflon (*O. musimon*) and Argali wild sheep (*O. ammon*) carried four copies of a 75 bp repeat, which discriminated them from Urial (*O. vignei*), and contained one 75 bp repeat followed by four 76 bp repeats (Table 1). The identity of each translation start and stop codon and the order of genes from the ovine mtDNA has been described previously (Hiendleder *et al.*, 1998). The order and features of the 16 mitogenomes were found to be the same, with the notable addition of a complete stop codon (TAA) in the *ND3* gene in two Urial sheep (*O. vignei* h75, h76). Following exclusion of the repeated motif in the control region, alignment of the mitogenome of domestic animals (16 264 bp) revealed 31 variations that were present in only one sequence and 306 parsimony informative sites. Domestic sheep nucleotide diversity (π) was calculated as $7.44 \pm 0.57 \times 10^{-3}$. Nuclear inserts of mtDNA (numts) are a widely accepted occurrence (Bravi *et al.*, 2006), however were unlikely to be present in this data set, given no heteroplasmic fragments were detected during either the PCR amplification or subsequent sequencing. Even so, the primers of Tapio *et al.* (2006); (Supplementary Table 1) were employed specifically to examine the samples for nuclear inserts of mtDNA signatures, but none were found.

Phylogenetic relationship between mitogenomes of domestic sheep

The sequence of all 16 mitogenomes (gen-mtDNA) was used to construct a phylogenetic tree to examine the relationship between domestic and wild sheep (Figure 1a). The mitogenomes from 10 domestic sheep clustered as expected, forming five groups that corresponded exactly with haplogroups HA, HB, HC, HD and HE. The average number of nucleotide differences and sequence diversity (*K*) were calculated to quantify the genetic distance separating haplogroups (Table 2). The greatest distance was observed between HB–HC ($D = 163.5$), closely followed by HB–HE and HC–HD ($D = 162.0$). The lowest number of observed differences was between HC–HE ($D = 58.5$) and HA–HB ($D = 93.0$), whereas all other domestic haplogroups were separated by more than 120 nucleotide differences.

Genetic distance between Urial, Argali and domestic sheep mitogenomes

The relationship between domestic and other *Ovis* spp. was resolved using the full mitogenome sequence. It is worth noting that the neighbour-joining tree constructed using the entire mitogenome resolved the position of each branch with 100% bootstrap support. The two European mouflon (*O. musimon*) sequences clustered together and were immediately adjacent to haplogroup HB. A total of only 11 substitutions were observed to distinguish the mouflon from the two domestic animals, which carried HB (Table 2). In contrast, over 350 mtDNA positions distinguished domestic sheep from the mitogenomes obtained from either Argali (*O. ammon*) or Urial (*O. vignei*) sheep ($D > 350$; $K > 2.0$; Table 2). The single Argali-derived mitogenome displayed higher divergence to domestic sheep sequences compared with any of the three Urial-derived sequences (Table 2). Comparison of the mitogenomes obtained from Urial and Argali sheep revealed greater than 370 substitutions (Table 2), indicating these two wild species are highly divergent.
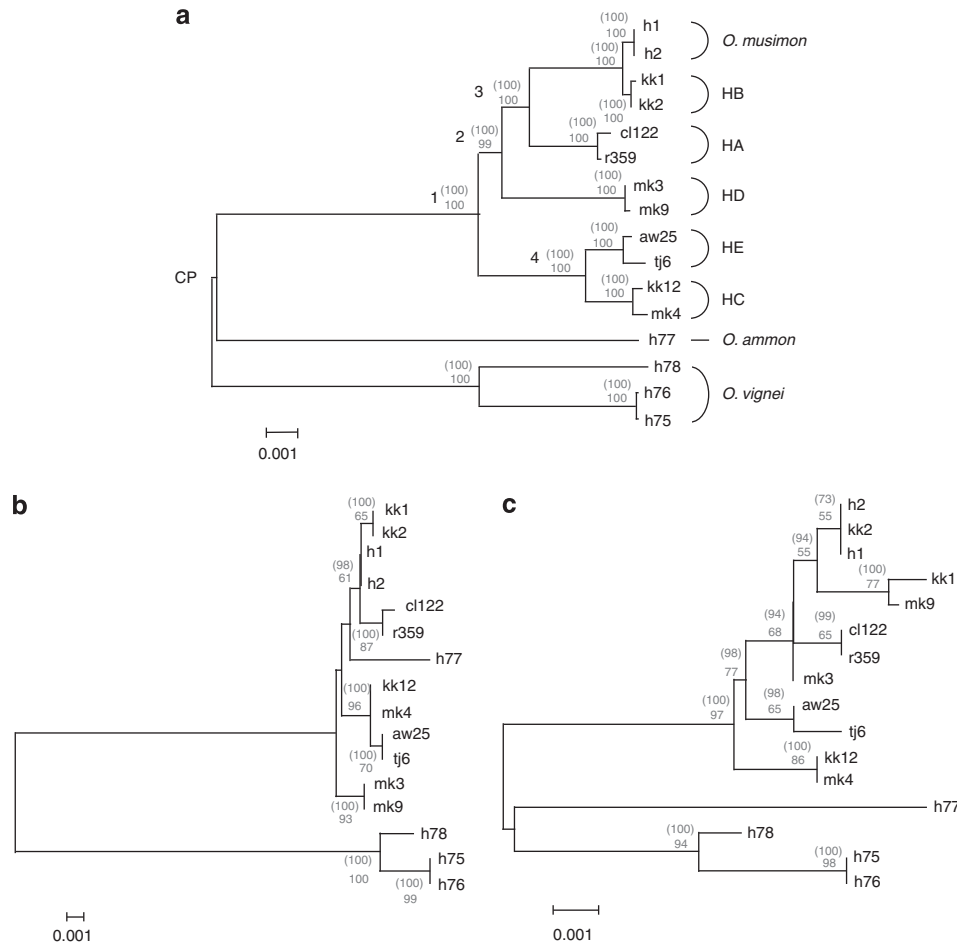
Figure 1 Phylogenetic relationship between six wild and ten domestic sheep inferred using the (**a**) complete mitochondrial genome sequence, (**b**) 12s rRNA and (**c**) tRNA concatemer. The neighbour joining trees were generated using Baysian and maximum likelihood methods with bootstrap support values are given on the nodes and clade credibility scores in brackets. The time since divergence of haplogroups A–E was calculated after first estimating the time since the most recent common ancestor of argali and domestic sheep. This was used as a calibration point (CP) for timing the radiation of each haplogroup A–E (nodes 1–4).

Table 2 Genetic diversity observed between domestic and wild sheep mitogenomes

	HA	HB	HC	HD	HE	Mouflon	Urial	Argali
HA	—	0.57	0.93	0.75	0.90	0.58	2.19	2.53
HB	93	—	1.01	0.81	1.00	0.07	2.31	2.59
HC	150.5	163.5	—	1.00	0.36	1.00	2.33	2.65
HD	122.5	131.5	162	—	0.98	0.81	2.27	2.61
HE	147	162	58.5	159.5	—	0.98	2.30	2.63
Mouflon	94	11	162.5	131.5	160	—	2.31	2.60
Urial	357.7	377	380.3	370.5	375.7	377.3	—	2.32
Argali	413	423	433	425.5	429	424	379	—

The average number of nucleotide differences (*D*) is given below the diagonal and nucleotide substitutions per site (*K*, given as a percentage) are given above the diagonal for the full mitochondrial sequence after removal of both indels and the repetitive component of the control region.

Phylogenetic reconstruction using mtDNA subsets

The availability of full-length mtDNA sequence provided the opportunity to examine whether the phylogenetic relationships observed in Figure 1a could be accurately recovered using subsets of the sequence. Maximum likelihood and Bayesian methods were used to generate 16 trees based on individual segments of the genome and three trees based on concatenated sequence from regions likely to be under similar evolutionary constraint (Figure 1b, c and Supplementary Figure 2). Out of 19 trees, 13 trees revealed a different branch structure compared with Figure 1a. Eight of the trees clustered the mouflon and HB sequences on the same branch as a result of the low number of substitutions that distinguished them. In addition, a number of haplogroup combinations were indistinguishable in trees based on *NDI* (HB and HC groups together), *COI* (HB, HC and HD groups together) and *ATPase8* (HA, HB and HD groups together) sequence (Supplementary Figure 2). The remaining discordant trees either placed the Argali sequence within the branch containing sequences from domestic sheep (12s rRNA, Figure 1b) or formed branches containing members drawn from different haplotypes (tRNA concatemer sequence, Figure 1c).

The unresolved and non-conforming phylogenies obtained using the 19 data subsets prompted analysis into which component of the mitogenome contributed most to the total genome tree presented in Figure 1a. Partitioned Bremmer support (PBS) analysis was used to derive a partial decay index for each partition by
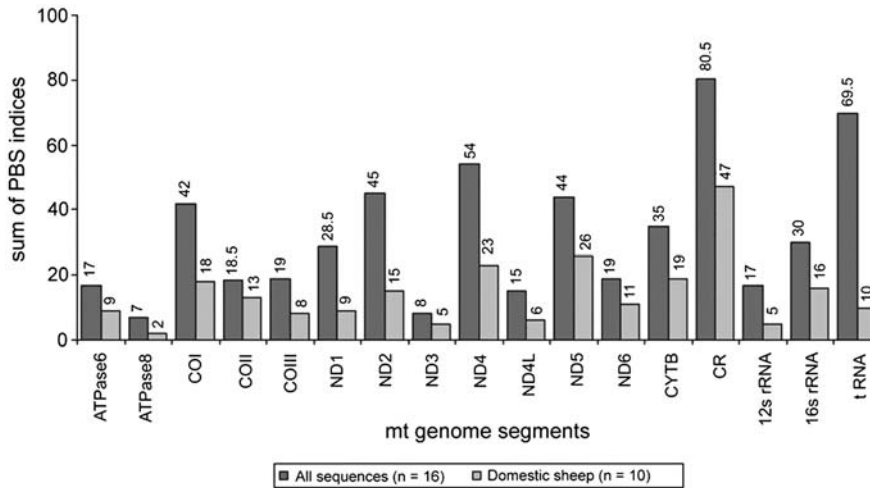
Figure 2 Partitioned Bremmer support (PBS) for different components of the mitogenome. The graph shows the relative contribution to the tree presented in Figure 1a of 17 components of the mitogenome. Analysis was performed either with (blue bars) or without (green bars) inclusion of the six animals classified as 'wild'. A full color version of this figure is available at the *Heredity* journal online.

subtracting the number of steps for that partition in the most parsimonious tree, from the shortest tree without the node in question (Sorenson and Franzosa, 2007). Given that each partition is mutually exclusive, the sum of partial indices for a given node is equivalent to the overall decay index. Figure 2 shows that the control region contributed the most to the total genome tree in both the presence ($80.5/549 = 14.7\%$) and absence ($47/242 = 19.4\%$) of wild sheep. Conversely, *ATPase8* sequence contributed the least (1.3 and 0.8%, respectively). The tRNA concatemer provided strong support for the genome tree, which included wild sheep (12.7%), reflecting the defined branching of *O. ammon* and *O. vignei* away from all other sheep (Figure 1c).

Dating the divergence of haplogroups A–E

Cytochrome B sequence was used to derive a calibration point relevant to the *Ovis* lineage given whole genome sequence is unavailable for closely related species. *Cyt B* data from the extinct *Myotragus*, thought to have diverged approximately 5.35 Myr ago, was used to date the split between Argali and domestic sheep to approximately 2.13 ± 0.29 Myr ago. This value was then used as a CP to estimate the divergence times separating the five haplogroups observed within domestic sheep using the concatenated sequence of all 12 protein-coding genes in the mitogenome. The oldest split (node 1, Figure 1a) was estimated to have occurred 0.92 ± 0.19 Myr ago. This separated the two most common haplogroups (HA and HB) from rarer haplogroups (HC and HE). Radiation of the two most commonly observed haplotypes (HA and HB) was estimated to have occurred 0.59 ± 0.17 Myr ago (node 3, Figure 1a), whereas the most recent domestic sheep divergence, HC vs HE, was dated to 0.26 ± 0.09 Myr ago.

Discussion

A major focus for the study was comparison of mtDNA sequence from both domestic and wild sheep. In previous studies, partial mtDNA sequences have been used to show the Urial and Argali are highly unlikely to be the progenitor of domestic sheep (Hiendleder et al., 1998, 2002; Wu et al., 2003; Chen et al., 2006; Tapio et al., 2006; Meadows et al., 2007), however the relationship between wild and domestic sheep was revisited in this study for three reasons. First, the relationship between wild species remains unresolved, as a number of different phylogenies have been reported (Hiendleder et al., 2002; Bunch et al., 2006; Meadows et al., 2007). Second, candidate wild species have been discounted as the progenitor of modern sheep using short fragments of mtDNA. Given wild and domestic sheep have overlapping ranges and can hybridise to produce fertile offspring, analysis of the full-length mtDNA may reveal additional information. Finally, additional and rare haplotypes have been identified in domestic sheep, which may represent a direct link with their wild ancestors (Guo et al., 2005; Pedrosa et al., 2005; Pereira et al., 2006; Tapio et al., 2006; Meadows et al., 2007). The neighbour-joining tree derived from the mitogenomes sampled this study (Figure 1) clearly demonstrated that neither Urial nor Agali sheep are the maternal ancestor of domestic sheep. This included domestic animals found to carry rare haplotypes belonging to haplogroups HC, HD and HE. In addition, Figure 1 revealed Argali sheep to be more closely related to domestic animals when compared with Urial sheep. This is inconsistent with a previous mtDNA-based analysis (Bunch et al., 2006), but concordant with a classification system based on chromosome number, which indicates that the Urial sheep ($2n = 58$) is more distantly related to domestic sheep ($2n = 54$) than the wild Argali ($2n = 56$) (Nadler et al., 1973). The mitogenome of the European mouflon (*O. musimon*) grouped with other haplogroup HB sequences in line with previous findings that indicate it should not be considered a truly wild sheep (Hiendleder et al., 2002; Pedrosa et al., 2005; Meadows and Kijas, 2008). Rather, the European mouflon likely represents a remnant from early domestication events which have readapted to a feral life (Chessa et al., 2009). The Asian mouflon, *O. orientalis*, is less well-characterised and thought to be the next closest extant *Ovis* species to domestic sheep based on *cytB* sequence (Bunch et al.,

2006; Tapio et al., 2006; Meadows et al., 2007). Unfortunately, the mitogenome of this moufloniform remains uncharacterised. If this was rectified, it would help to resolve the relationship between extant wild and domestic sheep and determine whether the wild progenitor is now extinct as is the case with the auroch and domestic cattle.

The sequence data was used to date the radiation of the haplogroups found in domestic sheep. This is a challenge as the fossil record of Ovis is poor, however the extinct M. balearicus was used as a CP in this study as it is thought to have been isolated around 5.35 Myr ago by the creation of the Strait of Gibraltar (Lalueza-Fox et al., 2005) and has an available mt sequence (cytB, AY380560). When Myotragus was used as the fixed age reference, the most recent common ancestor of O. ammon (h77) and other O. species was dated to 2.13 ± 0.29 Myr ago. It should be noted that only one CP was used, and the associated estimate is accompanied by a large confidence interval. Nevertheless, the estimate falls between previous estimates of 3.21–3.62 Myr (Hiendleder et al., 2002) and 0.8 Myr (Hernández Fernández and Vrba, 2005). The wide range of estimates reflects the poor fossil record and it should be recognised that the geological isolation CP for Myotragus–Ovis differentiation is expected to be a minimum age and the true time to most recent common ancestor between species may be much older. Having set the divergence time between Argali and domestic sheep to 2.13 Myr ago, the concatenated sequence of all 12 protein-coding genes was used to data the radiation of haplogroups HA–HE. The earliest radiation for haplogroup pairs detected in domestic sheep was placed at around 0.92 Myr ago. Comparison with other studies indicates the values reported here are generally lower than those obtained using the control region alone (Hiendleder et al., 2002) and higher than those obtained using cytB (Pedrosa et al., 2005). The haplogroup pair with the lowest sequence divergence was HC and HE ($K = 0.36$, Table 2). Given that relatively few animals have been identified, which carry HC and particularly HE, it is possible that these haplotypes represent sampling events at the extremes of a single large haplogroup. As a result, it is possible that the five haplogroups do not all represent independent lineages. The dating performed in this study estimated the time since divergence for HC–HE to be approximately 0.26 Myr. This is significantly older than archaeological evidence pointing to domestication around 8000–9000 years ago. It is also much older than available estimates for when population expansion occurred to speed the generation of new haplotypes within haplogroups (<20 000 years; Meadows et al., 2007). This suggests that the sequence divergence separating HC and HE is highly unlikely to be the product of expansion following domestication. Rather, both HC and HE likely represent the molecular signature of separate domestication events. It is, however, possible that both arose by a single domestication event that recruited highly divergent wild lineages. Taken together, this brings the minimum number of domestication events likely to have shaped the evolution of modern sheep to five, revealing a complex process involving a diverse ancestral population.

The utility of different components of the mitogenome for making phylogenetic inference was examined using data subdivision, tree construction and PBS indices.

Following removal of the hypervariable repeat, the control region displayed the largest PBS either with or without inclusion of the mitogenomes of wild animals (Figure 2). This is reassuring given that the control region has been used in almost every published study that aims to elucidate the maternal genetic history of sheep. Conversely, the analysis highlighted regions of the mitogenome, which should not be used in isolation to make phylogenetic inferences. The 12s rRNA displayed low PBS values and produced a tree with the least similarity to the full-mitogenome-derived tree (Figures 1c and 2). The COI gene is of interest following the proposal that it be used as the 'bar code' fragment for mammalian species designation based on genetic distance (Hebert et al., 2003). In this study, COI demonstrated higher than average PBS but failed to correctly define the HB–HC–HD–mouflon relationship. This indicated the gene would fail as a barcode across this relatively narrow Ovis evolutionary timescale. The identification of an optimal bar code is likely to become redundant as sequencing costs drop and the availability of next generation technologies make obtaining full mitogenomic sequences routine on a large scale. In this context, the 16 mitogenomes described in this work can be used as a reference set to classify new mitogenomes into established haplogroups, which have a fully-resolved phylogenetic relationship. In addition, the availability of this reference set may prove particularly useful for classification of short and fragmentary mtDNA sequence obtained from ancient and fossilised remains.

Conflict of interest

The authors declare no conflict of interest.

References

Bandelt H-J, Kong QP, Richards M, Villems R, Macaulay V (2006). Estimation of mutation rates and coalescence times: some caveats. In: Bandelt H-J, Macaulay V, Richards M (eds). *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag: Berlin. pp 149–179.

Bravi CM, Parson W, Bandelt H-J (2006). Numts revisited. In: Bandelt H-J, Macaulay V, Richards M (eds). *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag: Berlin. pp 31–46.

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

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**: 216–223.

Chessa R, Pereira F, Arnaud F, Amorium A, Goyache F et al. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**: 532–536.

706

Guo J, Du LX, Ma YH, Guan WJ, Li HB, Zhao QJ *et al.* (2005). A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim Genet* **36**: 331–336.

Hebert PD, Ratnasingham S, deWaard JR (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc London B* **270**: S96–S99.

Hernández Fernández M, Vrba ES (2005). A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biol Rev Camb Philos Soc* **80**: 269–302.

Hiendleder S, Lewalski H, Wassmuth R, Janke A (1998). The complete mitochondrial DNA sequence of the domestic sheep (*Ovis aries*) and comparison with the other major ovine haplotype. *J Mol Evol* **47**: 441–448.

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. *Proc Biol Sci* **269**: 893–904.

Lalueza-Fox C, Castresana C, Sampietro L, Marquès-Bonet T, Alcover JA, Bertranpetit J (2005). Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. *BMC Evol Biol* **5**: 70.

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

Meadows JRS, Cemal I, Karaca O, Gootwine E, Kijas JW (2007). Five ovine mitochondrial lineages identified from sheep breeds of the near East. *Genetics* **175**: 1371–1379.

Meadows JRS, Kijas JW (2008). Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. *Anim Genet* **40**: 119–123.

Moore S (1995). Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear gene trees. *Evolution* **49**: 718–726.

Nadler CF, Hoffman RS, Woolf A (1973). G-band patterns as chromosomal markers, and the interpretation of chromosomal evolution in sheep (*Ovis*). *Experientia* **29**: 117–119.

Pedrosa S, Uzun M, Arranz J-J, Gutierrezz-Gil B, San Primitivo F, Bayon Y (2005). Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proc R Soc B* **272**: 2211–2217.

Pereira F, Davis SJM, Pereira L, McEvoy B, Bradley DG, Amorim A (2006). Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Mol Biol Evol* **23**: 1420–1426.

Posada D, Crandall KA (1998). Modeltest: testing the modelof DNA substitution. *Bioinformatics* **14**: 817–818.

Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

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

Ryder ML (1984). Sheep. In: Mason IL (ed). *Evolution of Domesticated Animals*. Longman Group Limited: London and New York. pp 63–84.

Sanderson MJ (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol* **14**: 1218–1231.

Sanderson MJ (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**: 301–302.

Scherf BD (ed) (2000). *World Watch List for domestic animal diversity* 3rd edn Food and Agriculture Organization of the United Nations: Rome. p 58.

Sorenson MD, Franzosa EA (2007). *TreeRot, version 3*. Boston University: Boston, MA.

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

Tajima F (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**: 599–607.

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

Wood NJ, Phua SH (1996). Variation in the control region sequence of the sheep mitochondrial genome. *Anim Genet* **27**: 25–33.

Wu CH, Zhang YP, Bunch TD, Wang S, Wang W (2003). Mitochondrial control region sequence variation within the argali wild sheep (*Ovis ammon*): evolution and conservation relevance. *Mammalia* **67**: 109–118.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)