

## Molecular identification of sheep at Blydefontein Rock Shelter, South Africa

<sup>1</sup>K. Ann Horsburgh and <sup>2</sup>J. Víctor Moreno-Mayar

<sup>1</sup>Department of Anthropology, Southern Methodist University, Dallas, TX 75275; School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, P.O. Wits, 2050 South Africa; horsburgh@smu.edu

<sup>2</sup>Centre for GeoGenetics, Natural History Museum of Denmark, 1350 Copenhagen K, Denmark; morenomayar@gmail.com

### ABSTRACT

The Blydefontein Rock Shelter is a Later Stone Age archaeological site in the eastern Karoo of South Africa. No remains of domesticated animals have been reported although a dung layer, interpreted as deriving from sheep, dates to approximately one thousand years ago. The published morphological analyses of the site's fauna include many wild taxa, but also report that the majority of the bones in the assemblage were too fragmentary to identify. A recent re-examination of the assemblage identified ten specimens as examples of sheep or goats. In this paper we report on ancient DNA research on the eight specimens we were sent to study, six of which have preserved DNA. Of these, five are examples of wild animals, all of which had been previously identified as present at the site. One specimen was confirmed as a sheep, and it likely comes from a layer that dates to a period well after the initial introduction of domesticates. Direct dating of the specimen is not possible as the entire sample was consumed by the genetic testing. This study highlights the importance of ancient DNA as confirmation of taxon identification when the results of morphological identification challenge the broader culture history.

KEY WORDS: Faunal analysis, ancient DNA, Blydefontein, *Ovis aries*, *Capra hircus*.

Blydefontein Rock Shelter, in South Africa's eastern Karoo, has been excavated twice, once in 1967 and once in 1985 (see Figure 1 for a map showing the site's location, and Figure 2 for a stratigraphic section of the site). The faunal remains from those excavations were analyzed by Richard G. Klein and Kathryn Cruz-Urbe. The results were published, and at that time no domesticated fauna were reported (Sampson 1970; Klein 1979; Bousman 1998, 2005). Further, it was reported that the majority of the bones in the assemblage were too fragmentary to identify (Klein 1979: 36). In 2008 Ina Plug and Karin Scott undertook a re-analysis of the faunal remains recovered from the 1985 excavations and ten specimens were identified as domestic caprine (sheep or goat). One specimen in particular (BFT138) proved interesting as a consequence of a direct AMS (Accelerator Mass Spectrometry) date that seemed to provide evidence of domestic stock in southern Africa considerably earlier than the generally accepted date of approximately 2000 years ago (Sealy & Yates 1994; Pleurdeau et al. 2012).

Analyses of DNA from both modern (Horsburgh et al. 2013) and archaeological domestic fauna (Horsburgh & Rhines 2010; Orton et al. 2013) have proven valuable in reconstructing relationships among populations, and in verifying morphological identifications of skeletal elements (Loreille et al. 1997; Moss et al. 2006). There was some doubt about whether the caprine associated with the old radiocarbon date was from a sheep or a goat, so we were approached to attempt to recover ancient DNA (aDNA) from eight domestic caprine specimens (Bousman pers. comm. 2013).

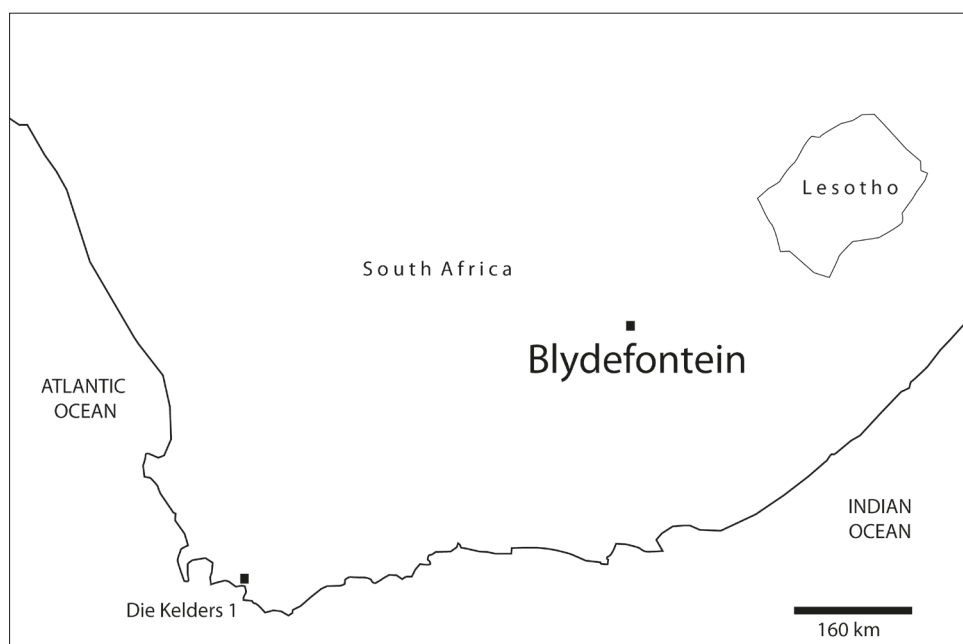


Fig. 1. Southern Africa, showing the location of Blydefontein Rock Shelter and Die Kelders 1. Redrawn after Bousman (2005).

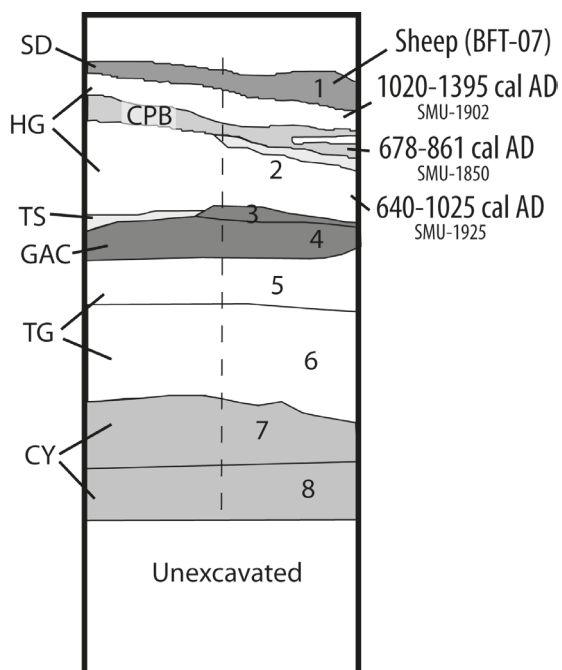


Fig. 2. Stratigraphic section of Blydefontein Rock Shelter showing the location of the genetically identified sheep specimen and relevant radiocarbon determinations. Redrawn after Bousman (2005).

### LABORATORY METHODS

All DNA extraction and Illumina library preparation before PCR amplification was undertaken in the University of Otago's purpose-built aDNA facility (Knapp et al. 2012). DNA was extracted from between 160 mg and 1.11 g of tooth or bone material using a standard silica and guanidinium thiocyanate protocol (Rohland & Hofreiter 2007a, b). Two negative controls were processed alongside the eight specimens.

Barcoded Illumina sequencing libraries were constructed directly from both the aDNA extracts and the associated negative controls using custom Illumina shotgun adapters as described by Meyer and Kircher (2010). The amplification plateau was estimated by quantitative PCR using SYBR Green dye (Applied Biosystems) on the Stratagene MxPro 3000P platform, and amplified libraries were visualized with SYBR Safe DNA Gel Stain (Invitrogen) on 2 % agarose gels. Libraries were immortalized by PCR amplification to plateau using ABI's AmpliTaq Gold with the following reagent concentrations: 1× AmpliTaq PCR Buffer, 2.5 mM MgCl<sub>2</sub>, 1 mM dNTPs, 0.2 μM of each extension primer and 3.75 units of AmpliTaq Gold. Immortalized libraries were purified using MinElute columns (QIAGEN) following the manufacturer's protocol with the addition of a second PE wash and a five-minute incubation with 0.1× TE buffer instead of the provided elution buffer.

Libraries were enriched for the mitochondrial genome by in-solution hybridization following Maricic et al. (2010), but employing a few modifications. Each sequencing library was enriched independently, and pooled in equimolar ratios only after a further 10 cycles of PCR amplification and quantification by qPCR as above. Additionally, the libraries were eluted from the MyOne Streptavidin C1 Dynabeads (Invitrogen) by heating for three minutes to 95°C instead of treatment with sodium hydroxide. The DNA used in the manufacture of sheep mitochondrial DNA capture bait was extracted from a lamb chop bought in a supermarket in Dunedin, New Zealand. Pooled libraries were sequenced on the Illumina MiSeq platform with 2 × 75 base paired-end reads.

### SEQUENCE PROCESSING AND MAPPING

First, we trimmed adapter-derived sequences, leading Ns and low-quality runs from the raw sequencing reads using AdapterRemoval 1.5.4 (Lindgreen 2012). We mapped the filtered reads to a set of reference mitochondrial genomes using the BWA-ALN algorithm version 0.7.5a-r405 (Li & Durbin 2009). Seeding (-l parameter) was disabled in order to prevent 5' terminal substitutions characteristic to aDNA to bias the mapping (Schubert et al. 2012). PCR duplicates were then identified and removed from resulting bam files using the *MarkDuplicates* command from the Picard tools suite (<http://broadinstitute.github.io/picard/>). We then performed indel-based local realignment on the reads using GATK (DePristo et al. 2011). Finally, the MD tag was recomputed for each read using the SAMtools 0.1.18 (Li et al. 2009) *calmd* command. Mapping statistics are reported in Table 1. The sequences are available in the supplementary online materials ([www.sahumanities.org](http://www.sahumanities.org)).

### SPECIES IDENTIFICATION FROM SEQUENCING READS

We took three complementary approaches for identifying the species to which the biological remains belong.

TABLE 1  
Mapping statistics. For each specimen the first three rows of statistics refer to mapping to *Onis aries*: GI:3445513, *Capra hircus*: GI:612342193 and *Bos taurus*: GI:662034268, and the last row refers to mapping to the assigned species.

Specimen	Total reads	Trimmed reads	# Mapped reads	# Mapped reads (w/o PCR duplicates)	% Endogenous (2 d.p.)	Mean read length	Coverage (2 d.p.)	Depth of coverage
BFT03	1146144	1139066	22629	3378	0.29	81.26	0.46	11.0
				3425	0.29	84.09	0.3	11.1
				18374	0.23	79.81	0.3	7.8
				116147	1.84	95.00	1.00	121.1
BFT07	260737	257604	7917	574	0.22	84.28	0.0	2.9
				131	0.05	79.07	0.26	0.5
				78	0.03	83.26	0.2	0.3
				572	0.22	84.00	0.91	2.9
BFT09	370584	366533	365	92	0.02	73.61	0.16	0.3
				82	0.02	69.88	0.12	0.3
				497	0.03	82.8	0.25	0.4
				NA	NA	NA	NA	NA
BFT133	1067987	1054310	8268	434	0.04	75.73	0.17	1.5
				8760	0.05	78.58	0.16	1.6
				6478	0.03	75.41	0.14	1.2
				39324	0.37	89.00	0.96	21.1

TABLE 1(*continued*)  
Mapping statistics. For each specimen the first three rows of statistics refer to mapping to *Onis aries*: GI:3445513, *Capra hircus*: GI:612342193 and *Bos taurus*: GI:662034268, and the last row refers to mapping to the assigned species.

Specimen	Total reads	Trimmed reads	# Mapped reads	# Mapped reads (w/o PCR duplicates)	% Endogenous (2 d.p.)	Mean read length	Coverage (2 d.p.)	Depth of coverage
BF-T134	660184	855415	899	395	0.04	76.22	0.2	1.4
			896	382	0.04	75.65	0.21	1.4
			606	264	0.03	74.14	0.16	0.9
			4589	2360	0.27	86.00	0.96	12.3
BF-T135	344958	341132	829	104	0.03	62.15	0.12	0.3
			713	91	0.03	64.32	0.09	0.3
			631	86	0.02	65.03	0.08	0.3
			NA	NA	NA	NA	NA	NA
BF-T137	660184	656024	2348	100	0.02	67.29	0.13	0.3
			2564	105	0.02	63.66	0.1	0.3
			1811	75	0.01	62.92	0.1	0.2
			8675	579	0.09	78.00	0.83	2.8
BF-T138	508668	504008	728	106	0.02	65.07	0.09	0.3
			734	109	0.02	58.79	0.09	0.3
			647	149	0.03	65.16	0.15	0.4
			722	332	0.07	66.00	0.54	1.3

### *Mapping-based approach*

We mapped the sequencing reads to three different reference mitochondrial genomes (*Ovis aries*: GI:3445513, *Capra hircus*: GI:612342193 and *Bos taurus*: GI:662034268) independently. We called majority-rule-based haploid consensus sequences using ANGSD (Korneliussen et al. 2014) after filtering reads with mapping quality lower than 30 and nucleotides with base quality lower than 20. We then performed a BLASTN (Altschul et al. 1997) search against the nucleotide database, using each consensus sequence as a query. For each of the three consensus sequences from each sample, we retrieved the best hit based on the e-value and coverage (Table 2).

### *De novo assembly-based approach*

Mapping-based consensus calling (and further species identification) is susceptible to being biased by the reference genome to which reads are mapped. Therefore, we also confirmed the results from the approach above through a *de novo* assembly procedure.

Since only a small fraction of the filtered reads appeared as endogenous mtDNA fragments (Table 1), we restricted this analysis to the reads that could be mapped to any of the mitochondrial references. For each sample, we assembled the corresponding reads using Velvet 1.2.03 (Zerbino & Birney 2008) with a k-mer length of 21 bases, and allowing for automatic coverage filtering. We used the resulting contigs as a query for a BLAST search against the nt database and retrieved the best hit for each contig. In Table 2 we report the species that was recovered by the majority of the contigs as a best hit, for each sample.

### *Species identification*

We assigned each sample to a candidate species based on the BLAST results from both approaches, which were highly concordant. For the mapping-based approach, we considered the results from the reference that yielded the most mapped reads, and for the assembly-based approach we considered the two most frequent BLAST hits. For the cases in which both approaches coincided in the genus assignment but not the species, the result from the assembly approach was kept.

We then mapped all the reads to a reference mitochondrial genome from the candidate species and generated consensus sequences, as described above. We then performed a BLAST search using such consensus sequences and assigned each sample to the best BLAST hit (Table 2). Note that samples for which we could not map 100 or more reads to at least two of the three reference genomes were not included at this stage.

## CYTOCHROME B ALIGNMENT

Cytochrome b (cyt b) is a mitochondrial gene involved in energy production in the cell. The DNA sequences of cyt b are relatively stable within species and variable between species so it has been one of the genes of choice for species identification in archaeology (Loreille et al. 1997), forensics (Parson et al. 2000; Lee et al. 2009) and conservation (Hsieh et al. 2001). Its reliability and efficacy for the purposes of species identification has been demonstrated across many taxa (Kocher et al. 1989; Irwin et al. 1991).

We aligned the cyt b sequences from each of the archaeological specimens with cyt b from sheep and goat, and all the members of the genus identified by the previous two approaches using Geneious 6.1.8 (Kearse et al. 2012). We list here the comparative

TABLE 2

Species diagnoses based on mapping available sequences to reference genomes and *de novo* assembly. Note that while BFT07 most closely resembles an *Ovis orientalis* genome deposited in GenBank (accession number KF938360.1), we have assigned BFT07 to *Ovis aries*, the domestic sheep, because *Ovis orientalis* (also known as *Ovis gmelinii* (IUCN/SSC Caprine Specialist Group 2000)) is the likely ancestor of modern domestic sheep (Demirci et al. 2013) and is therefore expected to be closely related to domestic sheep.

Specimen	Morphological identification	Morphological element	Mapping-based species identification	<i>De novo</i> assembly-based species identification	Cytochrome b comparison species identification	Species identification
BFT03	<i>Ovis aries</i>	Left M1	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>
BFT07	<i>Ovis/ Capra</i>	Complete 1 <sup>st</sup> incisor	<i>Ovis orientalis</i>	<i>Ovis orientalis</i>	<i>Ovis aries</i>	<i>Ovis aries</i>
BFT09	<i>Ovis/ Capra</i>	Proximal 2 <sup>nd</sup> phalanx	Excluded from analyses due to low DNA quality and quantity			
BFT133	<i>Ovis aries</i>	Calcaneum	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>	<i>Anidorcas marsupialis</i>
BFT134	<i>Ovis aries</i>	3 <sup>rd</sup> phalanx	<i>Pelea capreolus</i>	<i>Pelea capreolus</i>	<i>Pelea capreolus</i>	<i>Pelea capreolus</i>
BFT135	<i>Ovis/ Capra</i>		Excluded from analyses due to low DNA quality and quantity			
BFT137	<i>Ovis/ Capra</i>	Proximal 1 <sup>st</sup> phalanx	<i>Redunca fulvorifolia</i>	<i>Redunca fulvorifolia</i>	<i>Redunca fulvorifolia</i>	<i>Redunca fulvorifolia</i>
BFT138	<i>Ovis/ Capra</i>	Complete 3 <sup>rd</sup> phalanx	<i>Tragelaphus strepsaceros</i>	<i>Tragelaphus oryx</i>	<i>Tragelaphus oryx</i>	<i>Tragelaphus oryx</i>

species we used in each case, with the appropriate GenBank accession numbers. BFT137 was aligned with mountain reedbuck (*Redunca fulvorufula*, NC\_020742.1), southern reedbuck (*Redunca arundinium*, NC\_020794.1) and Bohar reedbuck (*Redunca redunca*, AF096626.1). BFT138 was aligned with eland (*Tragelaphus oryx*, NC\_020750.1), nyala (*Tragelaphus angasii*, NC\_020748.1), mountain nyala (*Tragelaphus buxtoni*, AY667216.1), bongo (*Tragelaphus eurycerus*, NC\_020749.1), lesser kudu (*Tragelaphus imberbis*, NC\_020619.1), bushbuck (*Tragelaphus scriptus*, JN632706.1), sitatunga (*Tragelaphus spekii*, NC\_020620.1), and greater kudu (*Tragelaphus strepsiceros*, NC\_020752.1). This sequence alignment allowed us to resolve the conflict between the species determinations derived from the mapping (kudu) and *de novo* assembly (eland) approaches and assign BFT138 to eland.

BFT03 and BFT133 were aligned with springbok (*Antidorcas marsupialis*, JN632596). Because the springbok is the only living member of its genus, we also aligned these sequences with those of the springbok's closest living relative (Bibi 2013), the gerenuk (*Litocranius walleri*, NC\_020716.1). BFT134 was aligned with the vaalribbok (*Pelea capreolus*, JN632684), which likewise is alone in its genus. We therefore aligned it with all the *Redunca* species listed above as well as all the members of the genus *Kobus*, specifically waterbuck (*Kobus ellipsiprymnus*, AF096625.1), kob (*Kobus kob*, AF052939.1), lechwe (*Kobus lechwe*, NC\_018603.1), Nile lechwe (*Kobus megaceros*, AF096620.1) and puku (*Kobus vardonii*, AF096619.1). In the interests of space we have not reproduced these sequence alignments, but they are available on request.

#### SPECIES IDENTIFICATION BY SEQUENCE DATA AND MORPHOLOGICAL DATA

Morphological analyses of archaeological fauna are integral to reconstructions of the ancient environments and the behaviour of the people living in them. They are, however, inherently subjective (Driver 1992; Wolverton 2013), and frequently much of the bone recovered from archaeological sites is unidentifiable to skeletal element or zoological taxon (Klein & Cruz-Urbe 1984). One of us has argued elsewhere that genetic data need to be interpreted carefully and integrated with other available lines of evidence, and that we should furthermore guard against the temptations to go beyond our data (Horsburgh 2015). We will, nonetheless, show here that the employment of genetic data to identify the species of fragmentary bones is considerably less fraught with potential misinterpretation.

The species at issue here are sheep (*Ovis aries*), goat (*Capra hircus*), springbok (*Antidorcas marsupialis*), eland (*Tragelaphus oryx*), mountain reedbuck (*Redunca fulvorufula*) and a vaalribbok (*Pelea capreolus*). In the interests of space we have not reproduced the complete ~16 500 base pairs of the mitochondrial genomes of each of the relevant species, although they are all available in the National Center for Biotechnology Information database, GenBank, maintained by the US National Institutes of Health ([www.ncbi.nlm.nih.gov/GenBank](http://www.ncbi.nlm.nih.gov/GenBank)). Figure S1 of the supplementary materials shows a phylogenetic tree constructed using a neighbour-joining method with a Jukes-Cantor model of molecular evolution executed in Geneious (Kearse et al. 2012). Genbank accession numbers of the specimens used are listed in the figure legend. The phylogeny visually represents the extent of evolutionary divergence among each of these species as reflected in the amount of DNA sequence divergence between them. Figure 3A shows DNA sequences from a small portion of cyt b from each of the relevant species



(base position numbers 14 159 to 14 392). Cyt b was chosen for this illustration because it has been widely used for the purposes of species identification elsewhere (Parson et al. 2000; Hsieh et al. 2001). Each of the DNA sequences presented in Figure 3A is identified by the species name, and each of the associated GenBank accession numbers is listed in the figure legend. In the presented example, the DNA sequences are aligned with the sheep mitochondrial reference genome. In the diagram the dots (.) represent positions in which a given sequence is identical to the sequence to which it has been aligned, in this case sheep. Where nucleotide bases are represented by letters in the sequence (A, C, T or G), the specified sequence differs from the sheep sequence to which it has been aligned. In this way the DNA variability between species can be observed. The degree of sequence difference is a reflection of the length of time to the most recent common ancestor, with mutations accumulating in the genome in a more or less clock-like fashion (Zuckerlandl & Pauling 1965).

Figure 3A shows the efficacy of using mitochondrial DNA sequences to identify species and to distinguish domesticated sheep and goats from wild bovid forms. Hence, we can have confidence in the validity of the genetic identification of the morphologically identified archaeological specimens.

Specimen BFT07 was morphologically assigned to sheep/goat. In this, the molecular results are consistent with the morphological diagnosis, and further, we are able to use the DNA sequence to distinguish between sheep and goat, showing that the specimen is from a sheep. Figure 3B shows the sheep and goat reference genomes aligned with BFT07. It can be seen that over the displayed bases, the goat reference sequence differs from the sheep reference sequence at 26 positions. In contrast, BFT07 does not differ at all from the sheep reference genome. When the entire mitochondrial genome of BFT07 is examined, it does differ from the sheep reference genome, as is to be expected; every species shows some degree of intraspecific genetic variation (Hsieh et al. 2001). By aligning 518bp of the mitochondrial d-loop of BFT07 with the d-loop sequences deriving from haplogroups A, B and C sequences reported by Guo et al. (2005), we were able to determine that BFT07 is a member of haplogroup B.

Figure 3C shows the two specimens we have identified as springbok (BFT03 and BFT133). To make the visual comparison straightforward we have aligned BFT03, BFT133, the sheep reference genome and the goat reference genome to the springbok genome. The dots in this case, then, indicate where the sequences are identical to the springbok mitochondrial genome. It can be seen that the two archaeological specimens are very much more similar to the springbok genome than they are to either the sheep or goat genomes. Figures 3D and 3E show the same types of alignment for specimens BFT134, genetically identified as vaalribbok (*Pelea capreolus*), and BFT137, identified as mountain reedbuck (*Redunca fulvorufula*).

The most critical specimen in these analyses is BFT138 because it is associated with a direct AMS date which, if it were a domestic specimen, would demonstrate a surprisingly early arrival of domesticated fauna in southern Africa. Figure 1F shows the recovered mitochondrial DNA sequence from BFT138 aligned with the eland (*Tragelaphus oryx*) mitochondrial genome above the homologous sequences from both sheep and goat. BFT138 is clearly an eland, not a sheep or a goat. Specimens BFT09 and BFT135 were poorly preserved and did not yield DNA of sufficient quality or quantity to allow determination of species.

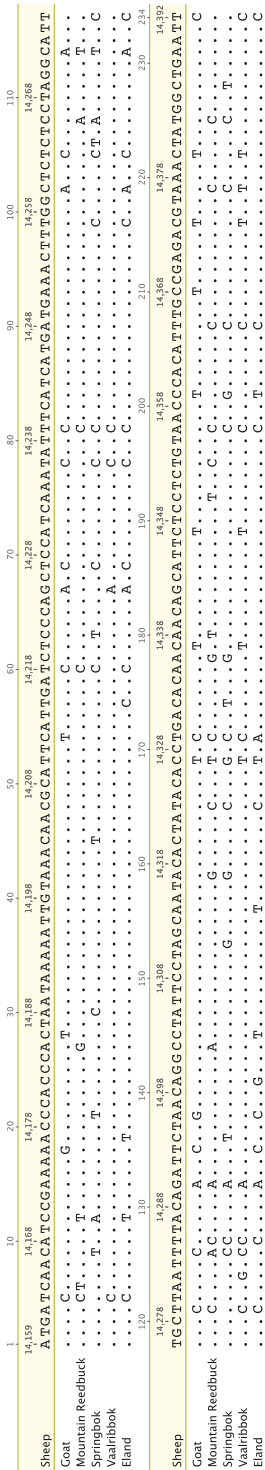


Fig. 3A. Sequence alignment of 234bp of the mitochondrial gene cytochrome b. In this figure, and subsequent figures, dots (.) represent nucleotide positions where the sequence for the given specimen is identical to the sequence at the top of the alignment. For example, in the first displayed position (nt 14,159), the sheep mitochondrial genome has an adenine residue (A). The other five shown species likewise have an A. All six species have the same nucleotides for the first four of the displayed positions (ATGA). At the fifth position, however, the sheep mitochondrial genome has a thymine (T), as does the springbok, but the other four species all have cytosines (C). It can be seen here that there are enough differences in the DNA sequences to confidently distinguish them from each other. The Genbank accession numbers for each of the DNA sequences displayed here are as follows: sheep (NC\_001941), goat (NC\_005044), mountain reedbuck (JN632695), springbok (JN632590), vaalibbok (JN632684) and eland (NC\_020750.1).

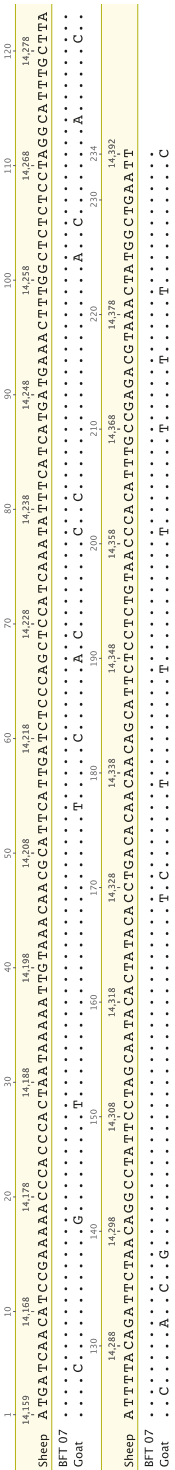


Fig. 3B. Sequence alignment of 234bp of the mitochondrial gene cytochrome b of sheep, goat and specimen BFT07. Over this stretch of the gene, the specimen BFT07 does not differ at all from the sheep reference genome, but differs at 26 positions from the goat reference genome.

1	10	20	30	40	50	60	70	80	90	100	110	120
Springbok	14,152	14,161	14,171	14,181	14,191	14,201	14,211	14,221	14,231	14,241	14,251	14,261
	A	T	G	A	T	C	A	A	A	A	A	A
BFT03	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
BFT 133	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Sheep	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Goat	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Springbok	14,281	14,291	14,301	14,311	14,321	14,331	14,341	14,351	14,361	14,371	14,381	14,385
	A	A	T	C	T	A	C	A	A	A	A	A
BFT03	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
BFT 133	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Sheep	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Goat	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

Fig. 3C. Sequence alignment of 234bp of the mitochondrial gene cytochrome b of sheep, goat and specimens BFT03 and BFT133 aligned to the springbok mitochondrial genome. The two archaeological specimens do not differ from the springbok reference genome at all, but differ from the sheep and goat genomes by 33bp and 40bp respectively.

1	10	20	30	40	50	60	70	80	90	100	110	120
Vaalribbok	14,455	14,154	14,164	14,174	14,184	14,194	14,204	14,214	14,224	14,234	14,244	14,254
	A	T	G	A	C	C	A	A	A	A	A	A
BFT 134	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Sheep	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Goat	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Vaalribbok	14,374	14,274	14,284	14,294	14,304	14,314	14,324	14,334	14,344	14,354	14,364	14,378
	T	A	G	T	C	T	A	C	A	A	A	A
BFT 134	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Sheep	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Goat	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

Fig. 3D. Sequence alignment of 234bp of the mitochondrial gene cytochrome b of sheep, goat and specimen BFT134 aligned to the vaalribbok mitochondrial genome. The archaeological specimen differs at only one position from the vaalribbok sequence, but differs from the sheep and goat sequences by 20bp and 18bp respectively.

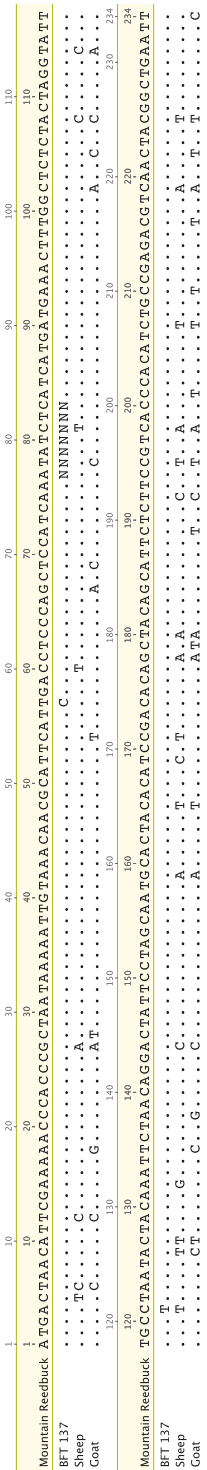


Fig. 3E. Sequence alignment of 234bp of the mitochondrial gene cytochrome b of sheep, goat and specimen BFT1137 aligned to the mountain reedbuck mitochondrial genome. The archaeological specimen differs at only two positions from the mountain reedbuck sequence, but differs from the sheep and goat sequences by 24bp and 36bp respectively. Note that the string of 7 Ns in the BFT1137 sequence indicates positions in the mitochondrial genome of the archaeological specimen that we were unable to recover.

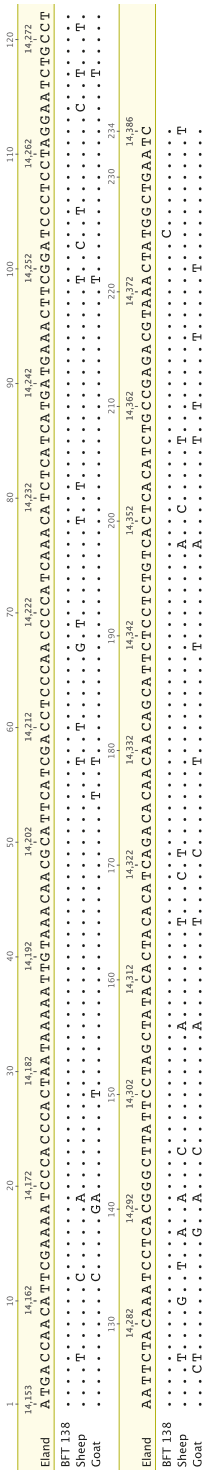


Fig. 3F. Sequence alignment of 234bp of the mitochondrial gene cytochrome b of sheep, goat and specimen BFT1138 aligned to the eland mitochondrial genome. The archaeological specimen differs at one position from the eland sequence, but differs from the sheep and goat sequences by 30bp and 23bp respectively.

### ANCIENT DNA DAMAGE PATTERNS

Based on the final sequence alignments to the assigned species, we investigated whether our sequences bore misincorporation patterns characteristic of aDNA, using *bamdamage* (Malaspinas et al. 2014). We observed an excess of C to T substitutions towards the 5' ends of the reads, which supports the authenticity of the extracted aDNA (Briggs et al. 2007; Krause et al. 2010; Skoglund et al. 2014). Note that in some cases, the expected complementary G to A substitutions towards the 3' ends are increased to a lesser extent. This is probably the consequence of reads being shorter than the actual DNA fragments that were built into sequencing libraries. The heavy tail present in read length distributions further supports this scenario (see Fig. S2 in the supplementary materials for plots of the base substitutions). We also observed an overall increased substitution rate, which could be attributed to the specificity limitations of our species assignment, which is ultimately limited by the species representation in the nucleotide database.

### NEGATIVE CONTROLS

We also produced sequencing data from two negative extraction controls, which we analyzed with the above mapping pipeline. As expected, these experiments yielded one and zero mapped reads (to any of the three reference mitochondrial genomes), respectively. This result, together with the observed misincorporation patterns towards the reads' termini, supports the authenticity of the genetic data produced in this study.

### DATING THE ARRIVAL OF DOMESTIC FAUNA

The specimen we were able to securely identify as sheep (BFT07) was excavated from Unit B8, Level 1. Before the specimens were made available for aDNA analysis, they had been subjected to AMS dating and stable isotope analysis. The remaining portion of specimen BFT07 was only 0.36 g of material, so it was therefore completely consumed by the DNA extraction protocols. There are no radiocarbon dates published from Level 1 of Blydefontein. There are, however, published dates from deeper in the sequence (Bousman 2005). We have taken those published, corrected, dates, and calibrated them using the SHCal13 calibration curve (Hogg et al. 2013). Level 2, dating to cal. AD 1020–1395 (SMU-1902) provides a lower boundary for the age of Level 1, as well as a date of cal. AD 640–1025 (SMU-1925) from a dung deposit (CPB) within Level 2, while a date of cal. AD 678–861 (SMU-1850) from just below the dung feature shows that the upper layers are well within the last 2000 years. The dung deposit has been previously interpreted as evidence for the presence of domestic stock (Bousman 1998, 2005), a finding consistent with our identification of a sheep bone in Level 1.

### DISCUSSION

A morphological re-analysis of the 1985 Blydefontein fauna identified ten specimens as either sheep or domestic caprine. This result differs from the original morphological analyses, which identified no domestic species. We had access to eight of the specimens identified as domestic fauna, and of those, six of the specimens' DNA preservation was sufficient to allow us to diagnose species confidently. Crucially, only one of those six specimens (BFT07) is a sheep. The remainder are wild species, all of which were identified by Klein and Cruz-Urbe in the original analysis (Bousman 1998).

All of Africa's domestic sheep originate outside the continent (Epstein 1971; Clutton-Brock 1993), most likely from several semi-independent populations in southwestern Asia (Meadows et al. 2011) where a dramatic decrease in the body size of caprines has been observed across the Aceramic Neolithic (Meadow 1984, 1993). There are few comparative data available for the mitochondrial genome of African sheep. A small fragment of the mitochondrial control region of twenty sheep specimens from the Later Stone Age deposits at Die Kelders 1 in South Africa's Western Cape Province has nevertheless been reported (Horsburgh & Rhines 2010). Like those from Die Kelders 1, the Blydefontein sheep is a member of haplogroup B, a clade with a now almost worldwide distribution (Hiendleder, Lewalski et al. 1998; Hiendleder, Mainz et al. 1998; Hiendleder et al. 2002; Demirci et al. 2013).

It is now evident from the mtDNA that the remaining five specimens represent wild species known to inhabit the region, including the specimen that yielded a surprisingly old direct AMS date, and previously argued to be a domestic species. The wild forms include an eland (*Tragelaphus oryx*), two springbok (*Antidorcas marsupialis*), a mountain reedbuck (*Redunca fulvorufula*) and a vaalribbok (*Pelea capreolus*). The original zooarchaeological analyses of the fauna recovered from the 1967 and the 1985 excavations at Blydefontein (Sampson 1970; Klein 1979; Bousman 1998, 2005) did not identify domesticates. Based on Klein's (1979) comment that the majority of the bones in the assemblage were too fragmentary to identify, we suggest it likely that the original analysts regarded the eight specimens under consideration here to be nonidentifiable on morphological grounds.

We do not suggest that aDNA analyses should be considered a replacement for traditional morphological analyses of archaeological fauna. Morphological analyses can be undertaken on much larger sample sizes than are practical for genetic analyses, which are tremendously costly in terms of both time and money. Furthermore, morphological studies provide information simply inaccessible to genetic analyses. In this instance, and others like it, aDNA analyses can provide a critical test of hypotheses of archaeological significance.

#### ACKNOWLEDGEMENTS

Horsburgh gratefully acknowledges salary support from a University of Otago Research Grant that also funded the laboratory costs. Moreno-Mayar was supported by the Lundbeck Foundation, the Danish National Research Foundation, and the Consejo Nacional de Ciencia y Tecnología (National Council of Science and Technology) in Mexico. We thank Britt Bousman for providing the specimens for analysis; Lisa Matisoo-Smith for the use of her laboratory; Zev Kronenberg for preliminary data processing; Richard G. Klein, David J. Meltzer and Jayson Orton for productive discussions and suggestions; and Mark D. McCoy for assistance in calibrating radiocarbon determinations and production of figures.

#### REFERENCES

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25** (17): 3389–402.
- Bibi, F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol Biol* **13**: 166.
- Bousman, C.B. 1998. The chronological evidence for the introduction of domestic stock into southern Africa. *African Archaeological Review* **15**: 133–50.
- Bousman, C.B. 2005. Coping with risk: Later Stone Age technological strategies at Blydefontein Rock Shelter, South Africa. *Journal of Anthropological Archaeology* **24** (3): 193–226.



- Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M. & Pääbo, S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104** (37): 14616–21.
- Clutton-Brock, J. 1993. The spread of domestic animals in Africa. In: T. Shaw, P. Sinclair, B. Andah & A. Okpoko, eds, *The archaeology of Africa: food, metals and towns*. London & New York: Routledge, pp. 61–70.
- Demirci, S., Bastanlar, E.K., Dagtas, N.D., Piskin, E., Engin, A., Ozer, F., Yuncu, E., Dogan, S.A. & Togan, I. 2013. Mitochondrial DNA diversity of modern, ancient and wild sheep (*Ovis gmelinii anatolica*) from Turkey: new insights on the evolutionary history of sheep. *PLoS One* **8** (12): e81952.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernysky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. & Daly, M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43** (5): 491–8.
- Driver, J.C. 1992. Identification, classification and zooarchaeology. *Circaea* **9**: 35–47.
- Epstein, H. 1971. *The origin of domestic animals of Africa*. Vol. 1. New York: Africana Publishing Corporation.
- Guo, J., Du, L.X., Ma, Y.H., Guan, W.J., Li, H.B., Zhao, Q.J., Li, X. & Rao, S.Q. 2005. A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics* **36** (4): 331–6.
- 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. *Journal of Molecular Evolution* **47** (4): 441–8.
- 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** (2): 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 B* **269** (1494): 893–904.
- Hogg, A.G., Hua, Q., Blackwell, P.G., Niu, M., Buck, C.E., Guilderson, T.P., Heaton, T.J., Palmer, J.G., Reimer, P.J., Reimer, R.W., Turney, C.S.M. & Zimmerman, S.R.H. 2013. SHCal13 Southern Hemisphere calibration, 0–50 000 years cal BP. *Radiocarbon* **55** (4): 1889–903.
- Horsburgh, K.A. 2015. Molecular anthropology: the judicial use of genetic data in archaeology. *Journal of Archaeological Science* **56**: 141–5.
- Horsburgh, K.A. & Rhines, A. 2010. Genetic characterization of an archaeological sheep assemblage from South Africa's Western Cape. *Journal of Archaeological Science* **37**: 2906–10.
- Horsburgh, K.A., Prost, S., Gosling, A., Stanton, J.A., Rand, C. & Matisoo-Smith, E.A. 2013. The genetic diversity of the Nguni breed of African cattle (*Bos* spp.): complete mitochondrial genomes of haplogroup T1. *PLoS One* **8** (8): e71956.
- Hsieh, H.M., Chiang, H.L., Tsai, L.C., Lai, S.Y., Huang, N.E., Linacre, A. & Lee, J.C. 2001. Cytochrome b gene for species identification of the conservation animals. *Forensic Science International* **122** (1): 7–18.
- Irwin, D.M., Kocher, T.D. & Wilson, A.C. 1991. Evolution of the cytochrome b gene of mammals. *Journal of Molecular Evolution* **32** (2): 128–44.
- IUCN/SSC Caprine Specialist Group. 2000. Workshop on Caprinae taxonomy, Ankara, Turkey, May 8–10.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. & Drummond, A. 2012. Gencious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28** (12): 1647–9.
- Klein, R.G. 1979. Paleoenvironmental and cultural implications of late Holocene archaeological faunas from the Orange Free State and north-central Cape Province, South Africa. *South African Archaeological Bulletin* **34**: 34–49.
- Klein, R.G. & Cruz-Uribe, K. 1984. *The analysis of animal bones from archaeological sites*. Chicago: The University of Chicago Press.
- Knapp, M., Clarke, A.C., Horsburgh, K.A. & Matisoo-Smith, E.A. 2012. Setting the stage: building and working in an ancient DNA laboratory. *Annals of Anatomy—Anatomischer Anzeiger* **194** (1): 3–6.
- Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X. & Wilson, A.C. 1989. Dynamics of mitochondrial-DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America* **86** (16): 6196–200.
- Korneliusen, T.S., Albrechtsen, A. & Nielsen, R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**: 356.

- Krause, J., Fu, Q.M., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P. & Pääbo, S. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464** (7290): 894–7.
- Lee, J.C.I., Tsai, L.C., Liao, S.P., Linacre, A. & Hsieh, H.M. 2009. Species identification using the cytochrome b gene of commercial turtle shells. *Forensic Science International: Genetics* **3** (2): 67–73.
- Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25** (14): 1754–60.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Proc. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25** (16): 2078–9.
- Lindgreen, S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* **5**: 337.
- Loreille, O., Vigne, J.D., Hardy, C., Callou, C., Treinen-Claustre, F., Dennebouy, N. & Monnerot, M. 1997. First distinction of sheep and goat archaeological bones by the means of their fossil mtDNA. *Journal of Archaeological Science* **24** (1): 33–7.
- Malaspinas, A.S., Tange, O., Moreno-Mayar, J.V., Rasmussen, M., DeGiorgio, M., Wang, Y., Valdiosera, C.E., Politis, G., Willerslev, E. & Nielsen, R. 2014. Bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics* **30** (20): 2962–4.
- Maricic, T., Whitten, M. & Pääbo, S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5** (11): e14004.
- Meadow, R.H. 1984. Animal domestication in the Middle East: a view from the Eastern margin. In: J. Clutton-Brock & C. Grigson, eds, *Animals and archaeology: 3. Early herders and their flocks*. Oxford: Oxford University Press, pp. 309–37.
- Meadow, R.H. 1993. Animal domestication in the Middle East: a revised view from the Eastern Margin. In: G. Posschl, ed., *Harappan Civilisation*. New Delhi: Oxford & IBH, pp. 295–320.
- Meadows, J.R.S., Hiendleder, S. & Kijas, J.W. 2011. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity* **106** (4): 700–6.
- Meyer, M. & Kircher, M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010** (6): pdb prot5448.
- Moss, M.L., Yang, D.Y., Newsome, S.D., Speller, C.F., McKechnie, I., McMillan, A.D., Losey, R.J. & Koch, P.L. 2006. Historical ecology and biogeography of North Pacific pinnipeds: isotopes and ancient DNA from three archaeological assemblages. *The Journal of Island and Coastal Archaeology* **1** (2): 165–90.
- Orton, J., Mitchell, P., Klein, R., Steele, T. & Horsburgh, K.A. 2013. An early date for cattle from Namaqualand, South Africa: implications for the origins of herding in southern Africa. *Antiquity* **87**: 108–20.
- Parson, W., Pegoraro, K., Niederstatter, H., Föger, M. & Steinlechner, M. 2000. Species identification by means of the cytochrome b gene. *International Journal of Legal Medicine* **114** (1–2): 23–8.
- Pleurdeau, D., Imalwa, E., Detroit, F., Lesur, J., Veldman, A., Bahain, J.J. & Marais, E. 2012. ‘Of sheep and men’: earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia). *PLoS One* **7** (7): e40340.
- Rohland, N. & Hofreiter, M. 2007a. Ancient DNA extraction from bones and teeth. *Nature Protocols* **2** (7): 1756–62.
- Rohland, N. & Hofreiter, M. 2007b. Comparison and optimization of ancient DNA extraction. *Biotechniques* **42** (3): 343–52.
- Sampson, C.G. 1970. *The Smithfield Industrial Complex: further field results*. Bloemfontein: National Museum.
- Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A., Willerslev, E., Krogh, A. & Orlando, L. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**: 178.
- Sealy, J. & Yates, R. 1994. The chronology of the introduction of pastoralism to the Cape, South Africa. *Antiquity* **68** (258): 58–67.
- Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J. & Jakobsson, M. 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **111** (6): 2229–34.
- Wolverton, S. 2013. Data quality in zooarchaeological faunal identification. *Journal of Archaeological Method and Theory* **20** (3): 381–96.
- Zerbino, D.R. & Birney, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18** (5): 821–9.
- Zuckerkandl, E. & Pauling, L. 1965. Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**: 357–66.