Quaternary Science Reviews 388 (2026) 110076

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

Evaluation of DNA and collagen preservation in Late Pleistocene and Holocene bovid fossils from South Africa

Deon de Jager [a,*], Andi M. Wilson [b,c], Alba Rey-Iglesia [a], J. Tyler Faith [d,e], Kaedan O'Brien [f], Wendy Black [g], Wilhelmina Seconna [g], Olivia Hall [h], Paul Szpak [h], Eline D. Lorenzen [a,*]

[a] Globe Institute, University of Copenhagen, Øster Voldgade 5-7, Copenhagen, 1350, Denmark
[b] Section for Organismal Biology, University of Copenhagen, Thorvaldsensvej 40, Frederiksberg, 1871, Denmark
[c] Forestry & Agricultural Biotechnology Institute, Department of Biochemistry, Genetics & Microbiology, University of Pretoria, Private Bag X20, Hatfield, Pretoria, 0028, South Africa
[d] Natural History Museum of Utah, University of Utah, 301 Wakara Way, Salt Lake City, 84108, USA
[e] Department of Anthropology and Archaeological Center, University of Utah, 260 South Central Campus Drive, Salt Lake City, 84112, USA
[f] Department of Anthropology, SUNY Oneonta, 108 Ravine Parkway, Oneonta, 13820, USA
[g] Iziko Museums of South Africa, 25 Queen Victoria Street, Cape Town, 8000, South Africa
[h] Department of Anthropology, Trent University, 1600 West Bank Dr., Peterborough, ON, K9L 0G2, Canada

ARTICLE INFO

Handling Editor: Dr P Rioual

Keywords:
Ancient DNA
Palaeoproteomics
Morphometrics
Radiocarbon
Africa
Palaeo-Agulhas Plain

ABSTRACT

Few palaeogenetic studies have been conducted in southern Africa, potentially due to the expectation of very poor biomolecular preservation and the lack of published exploratory studies evaluating this. Thus, we set out to test the preservation of DNA and collagen in skeletal remains over the last ~110,000 years in South Africa. We collected 320 Late Pleistocene and Holocene fossil teeth and bones from six bovid species, covering six fossil sites and generated digital records (3D models, photographs, and GBIF occurrences) before destructive sampling. We evaluated DNA preservation in 144 specimens and collagen in 54 of these. DNA was preserved in 45% and collagen in 35% of specimens. These were mostly Holocene in age (<11,700 years old), except for four Late Pleistocene specimens dating between ~50,000 and 12,000 years old harbouring authentic ancient DNA. Geological age was an important explanatory variable in DNA and collagen preservation, but a poor predictor of mean read length. In addition to age, fossil site was a significant explanatory variable in endogenous DNA content, and in the interaction between DNA and collagen preservation, which were also significantly correlated. Thus, collagen content could potentially be used as a predictor of DNA preservation, but only within individual sites. Single-stranded libraries incorporated up to 6.7-fold more endogenous DNA and had higher complexity than double-stranded libraries. We show that palaeogenetic studies on fauna at lower latitudes are possible, and we conclude with recommendations we deem important to consider to maximise the success of such studies.

1. Introduction

Ancient biomolecules, including DNA and collagen, can provide valuable insights into past biodiversity and ecology (A. Lindahl et al., 2025; Richter et al., 2022). They are commonly extracted from skeletal remains from archaeological and palaeontological contexts. We denote these as fossils, using a broad view of the term to indicate the preserved remains or traces of past life, in our case teeth and bones, that have not necessarily undergone the process of fossilisation, where organic material is replaced with inorganic material. DNA is preserved by its binding to both inorganic hydroxyapatite and organic collagen (Campos et al., 2012; T. Lindahl, 1993), while collagen is obtained from the organic fraction of fossil bones and teeth. Once extracted, ancient DNA can provide insights into the evolutionary history of extant and extinct species through palaeogenetic analyses, while collagen can provide the basis for radiocarbon dating, stable isotope analyses, and collagen fingerprinting.

There are numerous ancient biomolecular studies of faunas from the

* Corresponding authors
E-mail addresses: dejager4@gmail.com (D. de Jager), andi.wilson@plen.ku.dk (A.M. Wilson), alba@palaeome.org (A. Rey-Iglesia), jfaith@nhmu.utah.edu (J.T. Faith), kaedan.obrien@oneonta.edu (K. O'Brien), wblack@iziko.org.za (W. Black), wseconna@iziko.org.za (W. Seconna), oliviahall@trentu.ca (O. Hall), paulszpak@trentu.ca (P. Szpak), elinelorenzen@sund.ku.dk (E.D. Lorenzen).

https://doi.org/10.1016/j.quascirev.2026.110076
Received 10 March 2026; Received in revised form 12 May 2026; Accepted 20 May 2026
Available online 27 May 2026

Arctic and temperate regions, where permafrost and low average annual temperatures allow for better biomolecular preservation (Collins et al., 2002; Hofreiter et al., 2015). However, relatively few studies have examined wild faunas at lower latitudes, where higher temperatures generally result in poor DNA and collagen preservation. However, southern Africa played a role in the first ancient DNA study, when DNA was retrieved from the dried muscle and connective tissue of a piece of salt-preserved skin of a 140-year-old museum specimen of the extinct quagga (*Equus quagga quagga*), a subspecies of plains zebra endemic to South Africa (Higuchi et al., 1984).

In the four decades since, palaeogenetic data for wild southern African faunas have mainly been generated as an inadvertent by-product of archaeofaunal studies focused on domestic taxa from the late Holocene; the last ~3000 years (Horsburgh, 2008; Horsburgh et al., 2016; Horsburgh and Gosling, 2020; Horsburgh and Moreno-Mayar, 2015; Orton et al., 2013). Similarly, studies employing collagen for direct radiocarbon dating of wild faunal remains from the region are rare (but see Loftus et al. (2016)). Studies using collagen for taxon identification have typically focused on distinguishing domestic and wild faunas at archaeological sites, or on identifying the taxonomic origin of artefacts (Bradfield et al., 2019; Coutu et al., 2016, 2021; Janzen et al., 2021; Le Meillour et al., 2020, 2023). Stable isotope studies of wild faunas in the region typically target the better-preserved enamel from teeth for palaeodietary reconstruction (e.g. Sealy et al. (2020)).

Only two studies include older ancient genomic data from non-human faunas: the nuclear genome (0.9× depth-of-coverage) from the proximal phalanx (8.5% endogenous DNA) of a ~5800-year-old baboon, *Papio ursinus*, from Ha Makotoko in Lesotho (Mathieson et al., 2020), and the nuclear genome (2.1×) from the tooth root (3.5% endogenous DNA) of a ~9300-year-old extinct blue antelope, *Hippotragus leucophaeus*, from Nelson Bay Cave in South Africa (Hempel et al., 2022), a site situated along the southern coast of South Africa, which was formerly part of the Palaeo-Agulhas Plain.

The Palaeo-Agulhas Plain was an ecosystem on the southern coast of South Africa that was intermittently formed on the exposed continental shelf as sea levels fell during Quaternary glacial periods (Cleghorn et al., 2020; Fisher et al., 2010). It provided a large amount of additional habitat (~85,000 km^2 at its greatest extent), including grasslands, savannah floodplains, woodlands, wetlands, and fynbos, a wholly different ecosystem than the then-inland and present-day Greater Cape Floristic Region (Marean et al., 2020; Radloff et al., 2010). At the northern edge of the Plain, which is today's coastline, the foothills of the Cape Fold Mountains provided abundant rock shelters and caves where ancient humans and animals lived, preserving numerous archaeological and palaeontological sites that have contributed significantly to our understanding of past humans and ecosystems (Deacon, 1995; Faith et al., 2024; Henshilwood et al., 2002; Klein, 1972, 1974; Klein and Cruz-Uribe, 2016; Loftus et al., 2016; Mackay et al., 2014; Marean et al., 2014; Rector and Reed, 2010; Reynard and Wurz, 2020; Sealy et al., 2020).

Much of the Greater Cape Floristic Region has either a Mediterranean climate at present, with cool, wet winters and warm, dry summers, or experiences year-round rainfall. This may provide more favourable conditions for biomolecular preservation compared with the cool, dry winters and warm, wet summers of much of the interior and eastern parts of southern Africa, as heat and humidity both accelerate biomolecular degradation. Thus, with a strong investigative history, palaeoecological framework, high number of palaeo- and archaeological sites with high faunal diversity, and potentially favourable climate, the Greater Cape Floristic Region has excellent potential to be a regional hotspot of fossils with good biomolecular preservation.

To investigate the potential of fossils from sites in the Greater Cape Floristic Region for larger-scale biomolecular retrieval, we evaluated DNA and collagen preservation in the skeletal remains of six wild bovid taxa from six fossil sites in South Africa spanning the last ~110,000 years (Fig. 1). Given that fossil specimens are finite, we followed best-practice guidelines (Pálsdóttir et al., 2019) and made digital records of each specimen before any destructive sampling took place through photographs, 3D models, and recording fossil occurrences on the Global Biodiversity Information Facility. Thereafter, we investigated (i) the preservation of DNA in 144 specimens through time and across sites, and (ii) collagen preservation in 54 of these, given that the age of specimens from which ancient DNA is obtained is a crucial data point in palaeogenetic studies, and good collagen preservation is a proxy for the feasibility of direct radiocarbon dating (Wood, 2015). For specimens older than the radiocarbon limit (~50,000 years), collagen can still be used for stable isotope analyses and/or taxon identification. To our knowledge, this study is the largest palaeogenomic assessment of wild

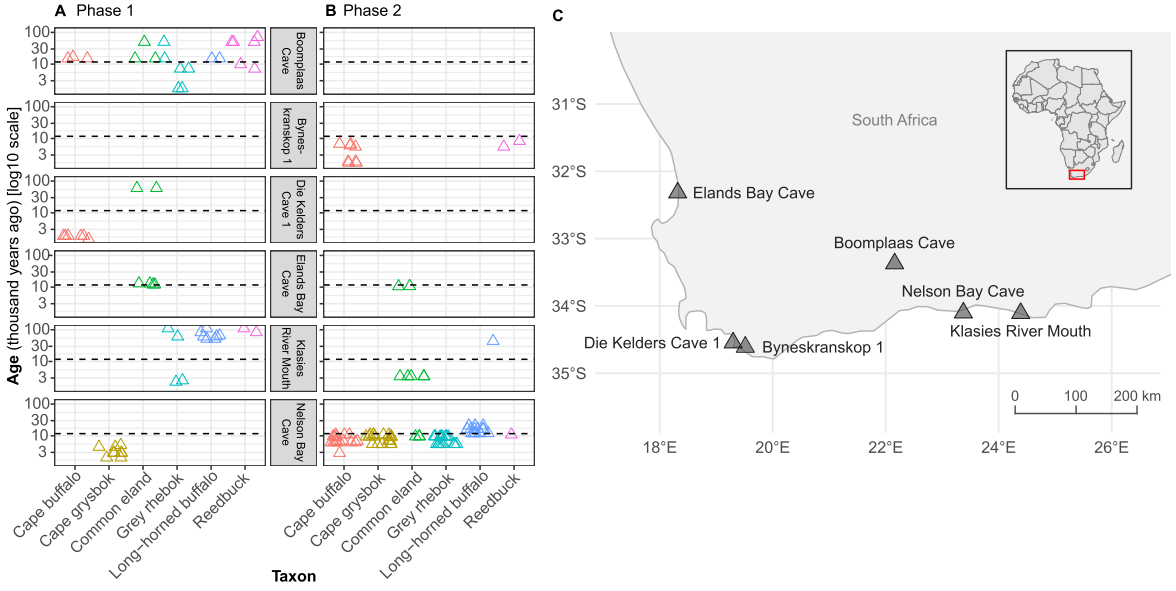

Fig. 1. Sampling scheme and sites investigated. **A** The 54 specimens collected in Phase 1 partitioned by taxon and site. **B** The 266 specimens collected in Phase 2 partitioned by taxon and site. The dashed lines in **A** and **B** indicate the Pleistocene-Holocene boundary (11,700 years ago). **C** Map showing the location of the six fossil sites. Note that Die Kelders Cave 1 and Byneskranskop 1 are very close geographically and thus their points were manually adjusted for clarity.

faunal remains in sub-Saharan Africa to date in terms of number of specimens screened. We therefore conclude by sharing considerations, based on our findings, that we deem important when undertaking such investigations in southern Africa that we hope will serve as a useful vantage point for others considering similar biomolecular work.

2. Materials and methods

2.1. Samples

A total of 320 available teeth (n = 304) and bone (n = 16) specimens from six bovid taxa spanning ~110,000 years in age were selected for analysis: Cape buffalo (*Syncerus caffer caffer*), the extinct long-horned buffalo (*Syncerus antiquus*), common eland (*Tragelaphus oryx*), grey rhebok (*Pelea capreolus*), Cape grysbok (*Raphicerus melanotis*), and reedbuck (*Redunca* spp.) (Fig. 1, Table S1). Bovids were targeted for this study as they are particularly abundant at fossil sites across southern Africa. The six taxa were targeted based on their fossil abundance and broad dietary category (grazer or browser), an aspect that will be explored in future work. Teeth were preferentially chosen over bone fragments, as we relied on previous identification of the specimens to species level (or genus level in the case of *Redunca*) and teeth can generally be more confidently assigned to species than bone fragments. All specimens were housed in the Archaeology Unit at the Iziko Museums of South Africa in Cape Town. For each specimen, a geological age range was estimated using stratigraphic inference based on published literature for each site from which the specimen originated (Table S1).

We targeted six fossil sites (Boomplaas Cave, Byneskranskop 1, Die Kelders Cave 1, Elands Bay Cave, Klasies River Mouth, and Nelson Bay Cave) related to the Paleao-Agulhas Plain in South Africa based on the abundance of bovid fossils in the museum collection and their importance in the study of past ecosystems and human-animal interactions in southern South Africa (see Introduction) (Fig. 1).

We sampled specimens in two phases. Phase 1 included the analysis of 54 specimens. These were used as proof-of-concept to determine whether DNA was preserved at all and if so, for how long. In late 2021, we selected specimens spanning the Late Pleistocene and Holocene (~110,000-1500 years ago) across five of the six sites (Fig. 1, Table S1). Based on the Phase 1 results, a report was evaluated by the South African Heritage Resources Agency (SAHRA), and Phase 2 was approved. This included the sampling of 266 additional specimens that were selected in mid-2022. Phase 2 sampling was focused predominantly on the Holocene and terminal Late Pleistocene (~44,000-1700 years ago), guided by Phase 1 results.

All specimens were exported to the Globe Institute, University of Copenhagen, Denmark, under SAHRA export permits 3348 (Phase 1) and 3884 (Phase 2) for destructive sub-sampling under SAHRA CaseID 16929 (Phase 1) and 20,048 (Phase 2). Destructive sub-sampling of the specimens was approved under a Section 35 permit from Heritage Western Cape (Case no.: 21081119SB0813E) and with documented permission from the curator (W. Black) and collections manager (W. Seconna) of the Archaeology Unit.

2.2. 3D models, photos, and occurrence data

Prior to destructive sub-sampling, a digital record was created of each of the 320 specimens via photography and 3D scanning. The latter was performed using either an Artec Space Spider or a 3Shape Trios dental scanner. The 3D models are available on MorphoSource (https://doi.org/10.17602/M2/L843966), while photos and a back-up of the 3D models are available on Zenodo (https://doi.org/10.5281/zenodo.19596154). The 3D models can be viewed, measured, and exported for 3D printing using the free, open-source software MeshLab: https://www.meshlab.net/ (Cignoni et al., 2008). The fossil occurrence data were also submitted to Global Biodiversity Information Facility (https://doi.org/10.15468/aafkx7).

2.3. Sub-sampling

All sub-sampling, DNA extraction, and library construction was conducted in a dedicated ancient DNA clean lab at the Globe Institute, University of Copenhagen, following strict ancient DNA protocols and practices. Negative controls were included for each DNA extraction and library construction batch to monitor for cross contamination and contamination with exogenous DNA.

All 54 Phase 1 specimens were screened for DNA and collagen preservation, regardless of putative age. Phase 2 samples were only assessed for DNA preservation, as only specimens with endogenous DNA will be considered for radiocarbon dating in a related study. Based on our Phase 1 findings of very poor DNA preservation in the older samples (~50,000 years or older), we focused our Phase 2 sampling on Holocene (<11,700 years ago) specimens. The exception was the long-horned buffalo specimens, where only specimens younger than 50,000 years old were screened as the species went extinct at the very end of the Pleistocene, ~11,700 years ago, in southern Africa and thus no Holocene specimens exist (Klein, 1984). Thus, while 266 specimens were initially exported for Phase 2, only 90 were screened for DNA preservation, though digital records were still generated for all specimens.

Phase 1 specimens (44 teeth, 10 bones) were manually processed in a clean lab, from sub-sampling to library construction. Phase 2 specimens (87 teeth, 3 bones) were manually sub-sampled, while DNA extraction and library construction were automated and performed on a Tecan Fluent 780 liquid handler (Tecan Group AG, Switzerland) at the GeoGenetics Sequencing Core, Globe Institute, University of Copenhagen. For a detailed description of the sub-sampling protocol for bones and teeth, and for general considerations when working with such specimens, see https://dx.doi.org/10.17504/protocols.io.q26g7n5o3lwz/v2 (de Jager et al., 2026).

2.4. Genetic laboratory methods

2.4.1. DNA extraction

DNA was extracted from powdered cementum or dentine (tooth) or bone material. For Phase 1 specimens, we used a silica spin column-based protocol, with the digestion buffer from Dabney et al. (2013) and the modified binding buffer from Allentoft et al. (2015). For details see the Supplementary information and https://dx.doi.org/10.17504/protocols.io.x54v9ronpv3e/v1 (Westbury et al., 2025). The optional pre-digestion was not performed. For Phase 2 specimens, an automated silica bead-based extraction described in McColl et al. (2024) was performed.

2.4.2. Library construction

Extracted DNA for both Phase 1 and 2 was treated with Thermolabile USER® II enzyme (New England Biolabs) at a final concentration of 0.15 U/μL for 3 h at 37 °C (aka "partial-USER", as no polynucleotide kinase was included), to remove most of the uracils at DNA fragment ends resulting from post-mortem cytosine deamination (Rey-Iglesia et al., 2026). Thereafter, single-stranded libraries were prepared using the Santa Cruz Reaction (SCR) (Kapp et al., 2021): https://dx.doi.org/10.17504/protocols.io.5jyl8d4p7g2w/v2 (Rey-Iglesia et al., 2025). This was done manually for Phase 1 samples and automated on the Tecan liquid handler for Phase 2 samples. Single-stranded (ssDNA) library methods are able to incorporate shorter DNA fragments than double-stranded (dsDNA) methods and are thus more suitable for samples with highly degraded DNA, as was expected to be the case with these specimens (Dabney et al., 2013; Kapp et al., 2021). Nonetheless, for the first batch of DNA extractions for the project, eight extractions representing six specimens, we also manually built double-stranded libraries using the Blunt-End-Single-Tube (BEST) protocol (Carøe et al., 2018): https://dx.doi.org/10.17504/protocols.io.4r3l21yq4g1y/v1 (Westbury et al., 2026) to compare to the ssDNA libraries from the same extractions (Table S3).

2.4.3. DNA sequencing

All libraries were dual-indexed using the uracil-tolerant polymerase, KAPA HiFi HotStart Uracil + ReadyMix (Roche), pooled in equimolar ratios and sequenced in eight separate sequencing pools. Phase 1 pools (PoolScreen01-07) were sequenced at Novogene UK on an Illumina NovaSeq 6000 using a 150 bp paired-end strategy. The Phase 2 pool (PoolScreen12) was sequenced at the GeoGenetics Sequencing Core on a NovaSeq 6000 using a 50 bp paired-end strategy. For more details, see the Supplementary information.

2.5. Collagen laboratory methods

Collagen preservation was evaluated only for Phase 1 specimens (n = 54). Collagen was extracted from tooth dentine (either root or crown) or bone powder at the Trent Environmental Archaeology Lab, Trent University, Canada, using a standard acid or acid-base-acid extraction including ultrafiltration using a 30 kDa filter. See the Supplementary information and Table S4 for more details.

2.6. Data analysis

2.6.1. Genetic species identification

Raw sequencing data was processed using PALEOMIX v1.3.9 (Schubert et al., 2014). Sequencing adapters were trimmed, overlapping reads merged, and reads shorter than 30 bp were discarded using AdapterRemoval v2.3.4 (Schubert et al., 2016). The clean, merged reads were aligned against a panel of 20 reference mitogenomes with BWA-backtrack/aln, BWA v0.7.18 (Li and Durbin, 2009), with seeding disabled (-l 1024), allowing a maximum of two gap opens (-o 2), and a relaxed edit distance threshold (-n 0.01) (Oliva et al., 2021; Schubert et al., 2012). Reads with a mapping quality <25 and duplicate reads were removed. The reference panel included the target species' reference mitogenome (except in the case of the extinct long-horned buffalo, for which no mitogenome is currently available), other co-occurring and/or closely related wild bovid species, three domestic species (cattle, sheep, and goat), and human (to detect any modern human DNA contamination) (Table S2). The number of unique reads aligned to each mitogenome was used to determine the genetic species ID. The mitogenome to which the highest number of reads aligned (if any) was assigned as the genetic species ID for that specimen, with a minimum of 10 reads aligned constituting a positive species ID. Thus, specimens were classified into two categories for whether the specimen was genetically identifiable: "Yes" and "No" (Fig. S1). In cases where the highest number of reads aligned to the human mitogenome (a result of contamination), the species was identified as the mitogenome with the next highest number of reads aligned. For specimens that were not genetically identifiable ("No") due to too few or no aligned reads, or equal number of reads aligned to multiple mitogenomes, the morphologically assigned species was retained as the species ID for the purposes of endogenous DNA estimation and plotting.

2.6.2. Endogenous DNA estimation

Following genetic species identification, the endogenous content of each library was calculated by dividing the number of unique reads aligned to the target nuclear genome by the total number of clean reads retained after AdapterRemoval filtering.

Clean, merged reads were competitively aligned (Feuerborn et al., 2020) to a concatenated reference genome containing the human GRCh38.p14 (GCF_000001405.40) and target species nuclear genomes using PALEOMIX as before, except that reads with a mapping quality <30 were discarded. Duplicate reads were removed. Additionally, microbe-like (Oskolkov et al., 2025) and mitochondrial-like regions were masked in the target nuclear genome. See the Supplementary information for more details.

The following nuclear reference genomes were used for the species indicated in parentheses after the genome: *Syncerus caffer* GCA_902825105.1 (Cape and long-horned buffalo samples); *Tragelaphus oryx* GCA_006416875.1 (common eland samples); and *Redunca redunca* GCA_006410935.1 (reedbuck and grey rhebok samples). For Cape grysbok samples, a species-specific reference assembly does not exist. Thus, we generated a fasta consensus sequence by mapping publicly-available reads (SRR37256366-SRR37256367) for this species to the sheep reference genome (Supplementary information). This was then used as the target nuclear reference genome for Cape grysbok specimens. Mitochondrial- and microbial-like regions were masked as described above.

2.6.3. Ancient DNA authentication

DamageProfiler v1.1 (Neukamm et al., 2021) was used to evaluate the ancient authenticity of endogenous reads by evaluating deamination rates and fragment lengths of mapped reads. Because reads were mapped to a concatenated reference of both human and target species nuclear genomes, the bam files were subset prior to DamageProfiler analysis to only include the target species scaffolds, and exclude mitochondrial- and microbe-like regions. This ensured that we were evaluating damage patterns of true endogenous reads and not including human, mitochondrial, or microbial reads.

2.6.4. DNA and collagen preservation

We evaluated DNA and collagen preservation by comparing the endogenous DNA content, the mean fragment length, and collagen yield against the age of specimens. Because specimen age ranges were stratigraphically-inferred, we conservatively used the lower (younger) age limit as the point estimate for the age of specimens.

Statistical analyses were performed using R and the *tidyverse* package v2.0.0 (Wickham et al., 2019). Four analyses were conducted: 1) to determine whether the age had an impact on the endogenous DNA content, 2) to determine whether the age had an impact on collagen yield, 3) to determine whether endogenous DNA content and collagen yield were correlated, and 4) to determine whether age had an impact on the mean read length. In analyses 1 and 3, the data representing the endogenous DNA content was log transformed. In all four analyses, univariate analyses were performed on the measured traits using linear or generalized linear models, depending on the distribution of each response variable. In addition to age, site, taxon, and material group (tooth/bone) were also tested as potential fixed effects in the models.

The significance of each of these factors on the measured trait was assessed using Type II ANOVA tests implemented in the ANOVA function of the *car* package v3.1-2 (Fox and Weisberg, 2019), after checking for normality, independence, and homoscedasticity, where relevant. Models with different fixed effects were compared using the corrected Akaike Information Criterion (AICc) calculated with the R package *performance* v0.12.2 (Lüdecke et al., 2021). F-tests were used for the Gaussian linear models and likelihood-ratio (LR) X^2 tests were used for generalized linear models.

2.6.5. Comparison of single-stranded and double-stranded libraries

We had constructed both SCR (single-stranded) and BEST (double-stranded) libraries from the same DNA extraction for six samples (see above, Table S3). Only two samples, RxA007 and ScA005, had endogenous DNA and were carried forward in this analysis. To compare the efficiency of these two library build methods, we equalised the number of reads by subsampling the raw data of both libraries for each sample to 9 million pairs using the reformat tool of the *bbmap* suite v39.01 (Bushnell, 2022). Reads from SCR and BEST libraries were processed identically and as described above in Endogenous DNA estimation and Ancient DNA authentication. We statistically compared the read length distributions between libraries within samples using a two-sample Kolmogorov-Smirnov test in R v4.4.1 (R Core Team, 2025). This tests whether two samples come from the same distribution and specifically determines whether or not the shape of the distributions are significantly different.

2.7. Results

2.7.1. Genetic species identification

In total, 144 specimens were screened for endogenous DNA, of which we were able to genetically identify 65 specimens (45%) to species level. No endogenous DNA was retrieved in the remaining 75 specimens (Fig. 1, Fig. S1, Tables S2–S3). The retrieved endogenous DNA reads showed typical ancient DNA damage patterns of increased C-to-T transitions at the ends as well as short lengths (Fig. S2). Thirteen negative controls were included across the eight sequencing pools and none contained reads from any of the target species, although five of these contained some reads that mapped to the human mitogenome (Fig. S1, Table S2). This indicates that there was no cross-contamination between samples during sub-sampling, DNA extraction, or library construction. The human contamination in the negative controls likely originated from persons performing the lab procedures. Three samples (RxA002, ScA001, ScA006) contained more reads mapping to the human mitogenome than the target species (Fig. S1, Table S2). In these cases, the human contamination could either originate from the lab procedures or from the handling of specimens without gloves during excavation and in the museum collection. Nevertheless, the human sequences were removed bioinformatically through competitive mapping and thus should not influence downstream analyses.

Of the 65 specimens genetically identified to species, 62 corresponded to the morphological identification made by palaeontologists. Two postcranial specimens tentatively attributed to reedbuck (*Redunca* spp.) in an unpublished analysis by the late James Brink, were genetically identified as grey rhebok (*Pelea capreolus*), a member of the same subfamily, Reduncinae (Fig. S1, Table S2). Both of these specimens, RxA007 (MuseumID: AL2 BLA N15 2.70) and RxA009 (AL2 BLA N15 2.72), were partial metatarsal bones from Boomplaas Cave. Another specimen, RxA002 (OCH MB O13 11-715), a partial molar tooth from Boomplaas Cave, was morphologically identified as southern reedbuck (*R. arundinum*), but was genetically identified as mountain reedbuck (*R. fulvorufula*) (Fig. S1, Table S2). This specimen has an estimated age of at least 50,000 years based on the stratum it is assigned to (OCH) from Boomplaas Cave (Faith, 2013; Miller et al., 1999; Vogel, 2001) (Table S1). While it does have some human DNA contamination (Fig. S1), the reads that mapped uniquely to the Bohor reedbuck (*R. redunca*) nuclear reference genome showed typical ancient DNA damage patterns and short read lengths, with a mean of 46 bp (Fig. S2). We additionally refined the identification of one reedbuck specimen to species level that was morphologically only identified to genus level (in Brink's unpublished analysis). This specimen, RxA008 (AL2A N15 2.132), is a complete 3rd phalanx bone from Boomplaas Cave that was genetically identified as a mountain reedbuck, *R. fulvorufula* (Fig. S1, Table S2).

In addition to RxA002, we retrieved endogenous DNA from three other Late Pleistocene specimens: PaA014 (YGL-D8-6), PaA015 (YGL-A6-1), and PaA018 (BSL-B5-4), which are long-horned buffalo, *Syncerus antiquus*, from Nelson Bay Cave. All three specimens are partial molar teeth (lower left m2, lower left m3, and upper right M3, respectively) (Fig. S1, Tables S1–S2). The age of PaA014 and PaA015 was estimated to be 23,500–21,000 years old, based on the lower and upper limits of layer YGL from the Bayesian modelled age chronology of Nelson Bay Cave by Loftus et al. (2016). The age of PaA018 was estimated to be 14,135–12,020 years old, based on two calibrated radiocarbon dates from layer BSL from Loftus et al. (2016). There was no human DNA contamination in the sequencing libraries of these specimens (Fig. S1) and the reads that mapped uniquely to the Cape buffalo nuclear reference genome showed typical ancient DNA damage patterns and short read lengths with means ranging between 43 and 44 bp (Fig. S2, Table S3).

2.7.2. Preservation of endogenous DNA and collagen

A total of 61/65 specimens with endogenous DNA were from the Holocene (<11,700 years old) with a mean age of 5742 years old, while only four were Late Pleistocene specimens (>11,700 years old) with a mean age of 26,005 years old (Fig. 2A). The percentage of endogenous DNA for genetically identifiable specimens ranged from 0.06% to 30% (Fig. 2A–Table S3). A total of 27/54 specimens evaluated for collagen preservation had an absolute collagen yield ≥1 mg, of which 20 were Holocene with a mean age of 3075 years old and seven were Late Pleistocene specimens with a mean age of 41,376 years old (Fig. 2B–Tables S3–S4). However, of these 27 specimens only 19 had a collagen yield of ≥1%, all of which were Holocene specimens with a mean age of 3128 years old (Tables S3–S4).

To determine whether age had a significant impact on endogenous DNA content (n = 144, Fig. 2A), we fit a linear model to the data, including $\log_{10}$(endogenous DNA content) as the response variable, and age and site as potential explanatory variables: lm (log_endo ~ age × site). This model fit the data better than models excluding site or including taxon or material group, based on the normality, independence, homoscedasticity, and AICc scores (Table S5, Figs. S3–S5). The model explained 42% of the variance in the endogenous DNA content (R^2 = 0.42, adj R^2 = 0.37). Age ($F_{1,132}$ = 5.75, $p < 0.05$) and site ($F_{5,132}$ = 4.67, $p < 0.001$) were found to significantly impact endogenous DNA content (Fig. S3). In general, an increase in age resulted in a decrease in endogenous DNA content and therefore DNA preservation. Notably, there was also a significant interaction between the age and site ($F_{5,132}$ = 7.94, $p < 0.001$). This indicates that the effect of age on DNA preservation differs between sites, suggesting that additional site-specific factors play a role in DNA preservation.

To determine whether age had a significant impact on collagen yield (n = 54, Fig. 2B, Figs. S3–S5), we fitted a generalized linear model to the data, using a gamma distribution with an inverse link. We included collagen yield (%) as the response variable and age as a potential explanatory variable: glm(collagen_yield ~ age, family = Gamma (link = "inverse")). This model fitted the data better than models including taxon or material group (Table S6). Here, site could not be tested as a fixed effect, as sampling was too uneven across sites, both for sample size and age range, to fit a generalized linear model, given that collagen data are only available for Phase 1 specimens (Fig. 1A). The model explained 52% of the deviance in collagen yield. Age had a significant effect on collagen yield (LR X^2 = 114.08, df = 1, $p < 0.001$). Overall, collagen yield decreased with increased specimen age.

To assess the relationship between endogenous DNA content and collagen yield (n = 54, Fig. 2C), we fitted a linear model to the data, including $\log_{10}$(endogenous DNA content) as the response variable and collagen yield and site as potential explanatory variables. This model fitted the data better than those excluding site or including taxon or material group (Table S7). The model explained 77% of the variance in the endogenous DNA content (R^2 = 0.77, adj R^2 = 0.72). Collagen yield ($F_{1,44}$ = 46.95, $p < 0.001$), site ($F_{4,44}$ = 7.14, $p < 0.001$), and the interaction between these two variables ($F_{4,44}$ = 2.79, $p < 0.05$) were all found to significantly impact endogenous DNA content. In general, an increase in collagen yield resulted in an increase in endogenous DNA content and therefore DNA preservation.

Finally, to determine whether age had a significant impact on read length (n = 65, Fig. 2D), we fitted a linear model to the data filtered for only genetically identifiable specimens. However, in this case the data were highly skewed in their distribution across sites. For example, 71% of the specimens were from Nelson Bay Cave (n = 46) while Elands Bay Cave only had one sample and Klasies River Mouth only two samples. Consequently, the analysis was restricted to only Nelson Bay Cave and a linear model including $\log_{10}$(mean read length) as the response variable and age as the potential explanatory variable was fitted to the data. This model fitted the data better than those including taxon or material group as fixed effects (Table S8). Age explained 19% of the variation in (R^2 = 0.19, adj R^2 = 0.17), and had a significant effect on mean read length ($F_{1,44}$ = 10.39, $p < 0.01$).

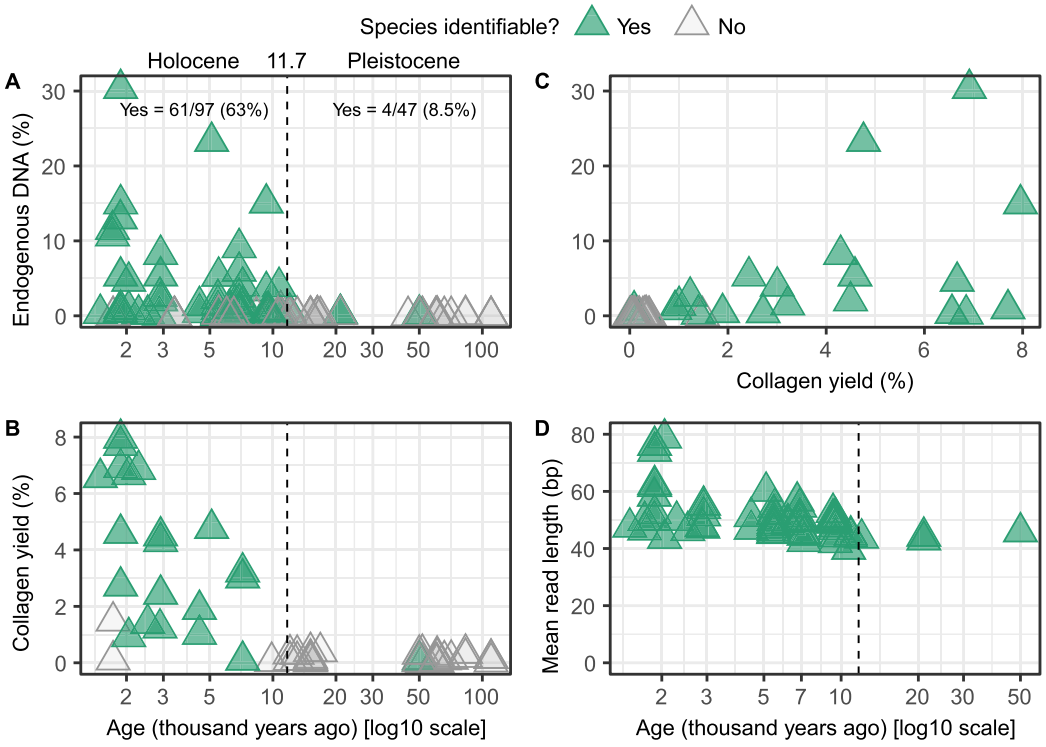

Fig. 2. Biomolecular preservation. A Endogenous DNA proportion as a function of specimen age for the whole dataset (n = 144). The text inside the plot indicates the proportion of specimens that were genetically identifiable ("Yes") in the Holocene and Pleistocene. **B** Collagen yield as a percentage of total dentine/bone powder, as a function of specimen age. **C** The relationship between DNA and collagen preservation, as represented by collagen yield in percentage and endogenous DNA proportion. Note that for **B** and **C**, only Phase 1 specimens (n = 54) are shown, for which we had both collagen and DNA data. **D** Mean length of endogenous reads as a function of specimen age (n = 65). Only genetically identifiable specimens are shown, as these have endogenous DNA from which to calculate read length. The vertical dashed line indicates the boundary between the Pleistocene and Holocene (11,700 years ago) in all plots. Species identifiable refers to whether the species could be genetically identified.

2.7.3. Comparison of SCR and BEST libraries

Starting with the same number of raw read pairs (9 million 150 bp PE) from the same DNA extraction, SCR libraries resulted in 4.9- and 5.4-fold more reads aligning to the target species reference genome than BEST libraries for RxA007 and ScA005, respectively (Table 1). This equated to 5.8- and 6.7-fold higher endogenous content in SCR libraries. The SCR libraries also had 0.6- and 2.4-fold lower duplication levels (clonality) than the BEST libraries. The read length distributions were significantly different between library methods ($D_{Kolmogorov\text{-}Smirnov}$ = 0.122 for RxA007, p = 0; and 0.161 for ScA005, p = 0), with SCR libraries producing a higher proportion of shorter reads compared to BEST libraries (Fig. 3).

2.8. Discussion

We investigated biomolecular preservation in Late Pleistocene and Holocene bovid fossils from sites associated with the Palaeo-Agulhas Plain in South Africa to determine whether, and over what timescales, DNA and collagen are preserved. While biomolecular studies have been done in southern African faunas, these have often focused on domestic taxa in Late Holocene archaeofaunal assemblages covering the past ~3000 years (Bradfield et al., 2019; Horsburgh and Gosling, 2020; Horsburgh and Moreno-Mayar, 2015; Le Meillour et al., 2023), or were focused on one fossil specimen of a wild taxon that preserved relatively high endogenous DNA (Hempel et al., 2022; Mathieson et al., 2020). We therefore present this analysis of biomolecular preservation in wild faunal fossils to provide a more comprehensive overview of what palaeogeneticists, palaeontologists, archaeologists, and other interested parties can expect when planning and conducting studies incorporating ancient DNA and collagen from southern African fossils.

We screened 144 fossils spanning the past ~110,000 years in age for DNA and 54 for collagen preservation, and found that biomolecular preservation was significantly impacted by estimated geological age of the specimen, as determined by stratigraphic context dating. This was not a surprising result, given that the post-mortem breakdown of DNA and proteins is known to be a partly time-dependent process (Adler et al., 2011; Allentoft et al., 2012; Hendy et al., 2018). We observed the notable pattern that 94% of fossils with sufficient endogenous DNA for species identification and 100% of those with a collagen yield of at least 1% (Dobberstein et al., 2009) were from the Holocene, thus less

Table 1
Comparison of SCR and BEST libraries. Note that overlapping read pairs were collapsed into single reads and only these collapsed reads were mapped to the reference genomes. M = million.

SampleID	LibraryID	Raw read pairs	Collapsed reads retained	Reads aligned to target	Endogenous proportion (%)	Clonality (%)	Mean read length (bp)	Standard deviation (bp)
RxA007	BEST_012	9 M	9.36 M	50,127	0.54	12.64	48.75	17.37
	SCR_002	9 M	7.96 M	245,634	3.09	8.15	44.46	14.34
ScA005	BEST_017	9 M	9.73 M	142,261	1.46	21.04	59.41	25.34
	SCR_007	9 M	7.89 M	774,843	9.82	8.95	50.97	19.30

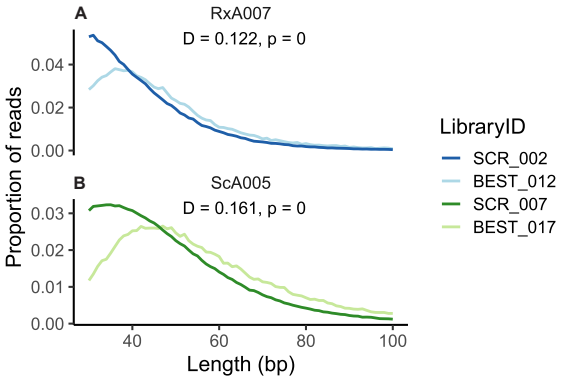

Fig. 3. Comparison of read length distributions of SCR and BEST libraries. The proportion of reads per length bin is shown for **A** RxA007 and **B** ScA005. The statistic reported is the two-way Kolmogorov-Smirnov test indicating that for both samples the read length distributions between SCR and BEST libraries differed significantly. Note that read length only up to 100 bp is shown here for visual clarity, due to the long tail of the distributions up to 289 bp.

than 11,700 years old (Fig. 1, Table S3). This indicates that it is unlikely, but not impossible, that these ancient biomolecules will be preserved beyond the Holocene. Indeed, we obtained endogenous DNA from four Late Pleistocene teeth ranging in age between ~50,000 and ~12,020 years old: RxA002 (age: ~50,000 years old; endogenous content: 0.2%), PaA014 (~21,000 years old; 0.3%), PaA015 (~21,000 years old; 0.3%), and PaA018 (~12,020 years old; 0.5%) (Fig. 1, Figs. S1–S2, Table S3).

2.8.1. DNA preservation

Ancient DNA preservation, measured as the proportion of endogenous DNA, decayed exponentially with age (Fig. 1), but was best modelled by taking both age and site into account (Fig. S3, Table S5). While the model including taxon also fit the data well, site and taxon were closely correlated. Thus, we focus on the best-fit model here, with age and site as explanatory variables. This model described 42% of the variation in endogenous DNA between specimens, where, in addition to age, site and the interaction between age and site were significantly associated with endogenous content. This indicates that site-specific factors also influence the preservation of DNA in fossil bones and teeth, and that the decay rate of DNA differs between sites.

Site-specific factors that may influence DNA preservation include soil pH, temperature, moisture and salt content, oxygen availability, exposure to UV radiation, and microbial activity (Campos et al., 2012; Hofreiter et al., 2015; T. Lindahl, 1993). Depurination, the predominant mechanism leading to strand breaks in DNA, occurs at lower rates at high or basic pH levels (Allentoft et al., 2012; T. Lindahl and Andersson, 1972), while lower temperatures reduces the rate of (bio)chemical reactions and microbial activity. Drier, saltier soil also leads to slower DNA degradation by limiting chemical reactions and microbial activity (T. Lindahl, 1993; T. Lindahl and Andersson, 1972).

The macro-characteristics of the site could also influence DNA preservation. For example, caves are generally cooler and have more stable temperatures than more open sites such as rock overhangs, due to less sun exposure, though all the sites included here are caves. In our case, it is difficult to determine which site best preserves DNA in skeletal fossils, as sites were not equally sampled through time, complicating comparisons of endogenous DNA from contemporaneous specimens (Fig. 1, Fig. S3). Some sites, such as Die Kelders Cave 1 and Klasies River Mouth, have two distinct human occupation periods resulting in the absence of bovid fossils for long periods (Schweitzer, 1979; Singer and Wymer, 1982). Others, such as Boomplaas Cave and Nelson Bay Cave, have an abundance of bovid fossil remains compared to Elands Bay Cave, for example. Thus, more specimens can be sampled from the former sites without having a significant impact on the collection. Two sites, Nelson Bay Cave and Byneskranskop 1, both have relatively consistent sampling across the Holocene to allow a like-for-like comparison in DNA preservation (Fig. 1, Fig. S3). A comparison of mean endogenous content of Holocene specimens between these sites revealed no significant differences (Wilcoxon signed-rank test for a difference in means: $p = 0.34$). Nonetheless, of the sites investigated, Nelson Bay Cave appears to have more favourable conditions than others, especially given that we obtained endogenous DNA from three Pleistocene specimens and Hempel et al. (2022) obtained a nuclear genome from an early Holocene blue antelope (*H. leucophaeus*) specimen from this site. However, in order to evaluate the factors affecting DNA preservation within and between Palaeo-Agulhas Plain sites in more depth, contemporaneous sampling of fossil remains would be needed in combination with measurements of soil characteristics, such as mean temperature, moisture content, and pH.

Additional factors that are difficult to account for *post-hoc* may further influence DNA preservation, such as post-excavation procedures and museum storage conditions. Pruvost et al. (2007) and Eriksen et al. (2025) found that freshly-excavated skeletal remains were a better source of ancient DNA than those stored in museums for 43-57 years, arguing that an excess of oxygen and temperature fluctuations accelerate DNA degradation post-excavation. In their study, Allentoft et al. (2012) found that storage time, one component of storage conditions, only had a small effect on DNA preservation. In combination, these studies imply that storage conditions other than time since excavation appear to be crucial for preserving any remaining endogenous DNA post-excavation. Thus, if ancient DNA work is planned for newly-excavated material, excavation procedures should include shielding fossils from direct sunlight to protect against UV radiation and heat, keeping fossils cold, and in as low-oxygen conditions as possible.

Having established that ancient DNA is preserved in South African fossils, we found that endogenous content ranged from 0.06% to 30%, with only 18% (26/144) of specimens having endogenous DNA above 1%. These were all of Holocene origin with a mean age of 4945 years old. Seven specimens had >10% endogenous DNA (Table S3), with a mean age of 3350 years old. The highest of these was 30% in a ~1875 year-old Cape buffalo tooth (ScA 004) from Die Kelders Cave 1. All seven were teeth, of which six were Cape buffalo from Die Kelders Cave 1, Nelson Bay Cave, and Byneskranskop 1. The seventh was a Nelson Bay Cave Cape grysbok specimen. The oldest of the seven, ScA029, was 9300 years old from Nelson Bay Cave and had an endogenous content of 15%. While the cut-off of 1% endogenous DNA mentioned above has no specific biological meaning, researchers may use this parameter as one of several criteria to decide which specimens warrant deeper sequencing without enrichment of the host DNA. If we consider 0.1% endogenous DNA, which Gretzinger et al. (2024) used as a cut-off to decide which libraries from ancient southern African human remains to enrich for human DNA via hybridization capture, we find that 41% (59/144) of specimens reached this threshold, with a mean age of 7114 years old (Table S3).

In addition to endogenous DNA content, another indicator of DNA preservation is the fragment length of recovered endogenous DNA. The mean read length across all genetically identifiable specimens in this study was 50 bp, ranging from 40 to 79 bp (Fig. 2). The best-fit model included only age as an explanatory variable for the variation in read length, but explained only 19% of this variation. It might be that DNA fragmentation occurs rapidly in the first few years after death (Campos et al., 2012) and then plateaus as the skeletal material starts to crystallise and the remaining DNA, protected by its binding to collagen and hydroxyapatite, fragments at a much slower rate (Allentoft et al., 2012).

In comparing the single-stranded SCR (Kapp et al., 2021) library build approach with the double-stranded BEST protocol (Carøe et al., 2018), we found the former incorporated significantly shorter fragments and produced more complex libraries with up to 6.7-fold more unique endogenous reads (Table 1, Fig. 3). Previous studies have shown

similar patterns and it appears that the advantages of single-stranded library methods are particularly pronounced when endogenous DNA content is below 3%; the median in our study was 0.7% (Table S3) (Gansauge et al., 2017; Kapp et al., 2021; Wales et al., 2015). Consequently, given the short read lengths and low endogenous DNA proportions observed in this study, it is crucial to employ DNA extraction (Allentoft et al., 2015; Dabney and Meyer, 2019; Rohland et al., 2018; Rohland and Hofreiter, 2007) and library build (Carøe et al., 2018; Gansauge et al., 2017, 2020; Kapp et al., 2021; Meyer and Kircher, 2010; Wales et al., 2015) methods that are able to recover and incorporate very short fragments when working with poorly preserved specimens.

2.8.2. Collagen preservation

Collagen preservation, expressed as collagen yield (%), was evaluated for the 54 Phase 1 specimens (Fig. 1), to investigate the feasibility of potential downstream analyses such as direct radiocarbon dating, stable isotope analyses for dietary reconstruction, and collagen fingerprinting for taxon identification. As with DNA, collagen preservation had a significant negative relationship with specimen age, where a generalized linear model incorporating only age explained 52% of the deviance in collagen yield (Fig. 2). A collagen yield of 1% is generally considered a reasonable cut-off for obtaining reliable elemental (C:N) and isotopic measurements (δ13C, δ15N), which in turn inform on the reliability of radiocarbon dating of the sample (Dobberstein et al., 2009). We found that 19 specimens had a collagen yield of 1% or more, ranging from 1.01% to 7.95%, all of which were of Holocene origin with a mean age of 3129 years old (Fig. 1). The specimen with the highest collagen content (7.95%) was ScA002, a ~1875 year-old Cape buffalo tooth from Die Kelders Cave 1, which had an endogenous DNA content of 15%.

While we could not include site as an explanatory variable in the generalized linear model due to unequal sampling through time between sites, we observed that collagen was preserved at four of the five Phase 1 sites. Elands Bay Cave was the only site where collagen was not preserved, though this was likely because specimens from this site were all between 10,500 and 13,100 years old, and not necessarily because it has poor conditions for collagen preservation (Sillen and Parkington, 1996). Nonetheless, should a more contemporaneous dataset of collagen measurements across sites be compiled, the contribution of site-specific environmental factors to collagen degradation could be investigated, as discussed for DNA above. In this case, alternative, minimally destructive methods of estimating collagen preservation could be used, such as attenuated total reflectance (ATR) Fourier Transform Infrared (FTIR) spectroscopy (Lebon et al., 2016), single-point near-infrared (NIR) spectroscopy (Ryder et al., 2026), or hot water-based extraction of soluble collagen (Higham et al., 2026).

A caveat in our study is that we used a 30 kDa filter during collagen extraction, as is standard protocol in preparing samples for radiocarbon dating. However, this method discards smaller protein fragments and thus selects for the best-preserved collagen samples. If the goal of collagen extraction was taxonomic identification by collagen fingerprinting, then either a smaller filter (e.g. 10 kDa) or no filter may be used (Wadsworth and Buckley, 2018). Collagen yield may then be higher and it might be that it remains preserved in older specimens than observed in our study.

2.8.3. DNA and collagen preservation relationship

Finally, we modelled the relationship between DNA and collagen preservation. Because some DNA is bound to collagen in skeletal remains (Campos et al., 2012), and evaluating collagen content via direct extraction, ATR-FTIR, or NIR is more cost effective than screening for endogenous DNA, measuring collagen content could potentially be used as a screening tool to determine which specimens to target for ancient DNA extraction. Previous studies have found variable results when evaluating the relationship between DNA and collagen preservation in skeletal remains. Ottoni et al. (2009) did not find a correlation between DNA preservation and intact collagen fibrils and obtained more DNA from specimens with damaged collagen from burning. Schwarz et al. (2009) also did not observe a correlation in very well-preserved mammoth (*Mammuthus primigenius*) remains, which Campos et al. (2012) propose is due to the excellent preservation, as the latter study only found a positive correlation in musk ox (*Ovibos moschatus*) specimens with low amounts of DNA, thus relatively poorer preserved specimens. Several other studies have also found a positive correlation between the preservation of these biomolecules (Campos et al., 2012; Poinar et al., 1996; Poinar and Stankiewicz, 1999; Sosa et al., 2013), which our findings also support.

While collagen yield alone explained 56% of the variation in endogenous DNA content, the model fit was relatively poor compared to when site was included as an additional explanatory variable (Table S3). The latter model explained 77% of the variance in DNA preservation. This correlation appears to be one of the strongest yet published, which may be because the previous studies relied on PCR amplification, which often included amplification of relatively large fragments of ancient DNA (e.g. 62-912 bp) (Ottoni et al., 2009; Poinar and Stankiewicz, 1999; Schwarz et al., 2009). Here, we employed lab techniques optimized for retaining very short DNA fragments and shotgun sequencing, and thus likely obtained a more accurate representation of the endogenous DNA content of the specimens.

Nonetheless, our results indicate the relationship between the preservation of these two biomolecules is complex, and that collagen yield is only partially predictive of endogenous DNA content, together with site. Thus, while broad predictions of DNA preservation across several sites in a region based on collagen preservation might not be highly informative, predictions within a single site might be useful. It should be noted again that our approach of using a 30 kDa filter may also have obscured some of the correlation between collagen and DNA preservation, and collagen yields from extractions without a filter or a smaller filter, or different methods of measuring collagen/protein content might result in an even stronger correlation. Indeed, Sosa et al. (2013) found a significant relationship between the presence of ancient DNA (qualitative: yes/no) and the infrared splitting factor from ATR-FTIR measurement of bone protein content, but whether this relationship holds with a quantitative measurement of DNA preservation remains to be investigated. Furthermore, DNA is not only associated with collagen in skeletal remains, but is also bound to the inorganic hydroxyapatite fraction (Campos et al., 2012; Götherström et al., 2002), and DNA and collagen do not necessarily follow the same decay kinetics (Dobberstein et al., 2009). Thus, the relationship between DNA and collagen preservation in skeletal remains will always be imperfect, and the absence of collagen does not necessarily mean DNA is not preserved or *vice versa*.

2.8.4. Considerations for ancient DNA faunal studies in southern africa

Predicting biomolecular preservation in skeletal remains is a difficult task. While latitude, a broad proxy for temperature, and geological age of the specimens can be used as a rough guide, local conditions of the site and soil surrounding the specimen are also important factors affecting biomolecular preservation. These include UV exposure, soil pH, moisture, and salt content, as well as microscopic processes such as decomposition by microbes and the interaction between DNA and organic and inorganic molecules during diagenesis (Campos et al., 2012; Dobberstein et al., 2009). Furthermore, post-excavation procedures and storage conditions in the field and museums can also negatively affect DNA preservation (Eriksen et al., 2025; Pruvost et al., 2007). As a result, and due to the fact that fossil remains are finite (Pálsdóttir et al., 2019), as are resources, every effort should be made to maximise the chances of obtaining endogenous DNA from such specimens. Thus, we present a short list of aspects from specimen selection (also see https://dx.doi.org/10.17504/protocols.io.q26g7n5o3lwz/v1) through to library building we deem important to consider to improve the probability of success of palaeogenetic projects, and specifically those on southern African faunas.

1. Make physical and/or digital copies of specimens before subsampling, through casting, 3D scanning, and/or photography, for the benefit of future studies. Care should nevertheless still be taken to be as minimally destructive as possible during subsampling.
2. Targeting Holocene specimens will provide the best chance of success.
3. However, for special-interest cases where Pleistocene specimens are to be targeted, such as species that went extinct in the Pleistocene, researchers can maximise their chances of success by following the guidelines below. Additionally, sites that have previously been shown to have good DNA preservation (e.g. Nelson Bay Cave, Byneskranskop 1) could be targeted. In cases where biomolecular preservation at a site has not been evaluated, teeth with preserved cementum could be targeted.
4. As DNA continues to degrade while specimens are in storage in museum collections, freshly or more recently excavated specimens should be targeted whenever possible, especially in the case of Pleistocene specimens. Though note that all specimens in this study were stored in museum collections for decades, as are most available specimens, and thus accessioned material is still a highly valuable resource in ancient DNA work. Nonetheless, for improved chances of success, ancient DNA work should be considered in excavation project design from the start, especially so that specimens targeted for ancient DNA analysis can be handled appropriately in the field (e.g.: with gloves, kept cold and out of the sun).
5. For sub-sampling of specimens, target the cementum of tooth roots (the outer layer), which generally better preserves endogenous DNA compared to dentine, as shown in humans (Damgaard et al., 2015). This has the advantage that the same root can then be targeted for collagen via the dentine that is left behind (if the tooth is big enough).
6. While only a few bones were screened in this study, DNA preservation appeared to be worse in bones than in tooth roots (Fig. S3). If bones must be sub-sampled, target the dense/hard outer layer, which contains the highest proportion of endogenous DNA, as shown in Pleistocene cave bear (*Ursus spelaeus*) and leopard (*Panthera pardus*) specimens (Alberti et al., 2018).
7. While we used a destructive sub-sampling method (drilling) for our DNA and collagen extractions, the fact that the outer layer of tooth roots and bones often harbour the highest concentration of endogenous DNA suggests minimally destructive methods of biomolecular extraction should be feasible and effective on southern African fossil remains, and we encourage future studies to explore this (Essel et al., 2021, 2023; Higham et al., 2026; McGrath et al., 2019; Richter et al., 2022; Scarsbrook et al., 2023; van Doorn et al., 2011). However, it should be noted that minimally-destructive extraction of either DNA or collagen might deplete the other in the specimen, especially if the entire specimen is used for this purpose (Higham et al., 2026). Thus, researchers should carefully consider this approach and how it might affect potential downstream analyses.
8. Because the outer layer of both tooth roots and bones seem to contain the highest proportion of endogenous DNA, do not remove this outer layer for the purposes of decontamination. Instead, wipe with 5% bleach followed by 70% ethanol or water to remove the bleach before sub-sampling.
9. Build single-stranded DNA libraries, as these incorporate much shorter fragments of DNA than blunt-end double-stranded DNA libraries. In southern Africa, this is very important, as the preservation is generally quite poor and so the ancient DNA fragments are very short and the endogenous content is low. Thus, researchers must maximise the number of fragments incorporated into the library.
10. Due to poor DNA preservation, palaeogenetic studies in southern Africa will most likely require enrichment of the sequencing libraries for host DNA via hybridization capture, both for mitochondrial and nuclear genomes. Thus, researchers should plan for this in project design and budgeting.

2.9. Conclusion

In this study, we presented an evaluation of biomolecular preservation in wild faunal remains from South African fossil sites. We did this to provide the community working with fossils from southern Africa who might be interested in using DNA and collagen to answer various evolutionary, archaeological, and palaeontological questions with an overview of how well and for how long these molecules might be preserved. We hope this will facilitate researchers in their study designs and research question formulations to allow for realistic approaches and efficient use of limited material and resources. Furthermore, we showed that DNA and collagen are not only preserved in late Holocene, but also in mid-to early Holocene specimens, and even Late Pleistocene specimens in rare cases. We believe that this will encourage more biomolecular research into the ancient wild faunas of southern Africa to expand our knowledge about their evolutionary history.

Ethics approval statement

All specimens were exported to the Globe Institute, University of Copenhagen, Denmark, under SAHRA export permits 3348 (Phase 1) and 3884 (Phase 2) for destructive sub-sampling under SAHRA CaseID 16929 (Phase 1) and 20,048 (Phase 2). Destructive sub-sampling of the specimens was approved under a Section 35 permit from Heritage Western Cape (Case no.: 21081119SB0813E) and with documented permission from the curator (W. Black) and collections manager (W. Seconna) of the Archaeology Unit.

Data accessibility and benefit-sharing

All raw sequencing data are available on the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB107966 (https://www.ebi.ac.uk/ena/browser/view/PRJEB107966). Metadata are also stored in the ENA (BioProject PRJEB107966) using the Host Associated MIxS package ERC000013 with manual addition of metadata terms from v0.8.1 of the MInAS checklist (Minimum Information about any Ancient Sequence) (https://www.mixs-minas.org/extension-ancient/). Code for analyses and plots is available on GitHub at: https://github.com/DeondeJager/biomolecular-preservation-south-africa. Specimen occurrence records are available on GBIF: https://doi.org/10.15468/aafkx7. 3D models of specimens are available on MorphoSource: https://doi.org/10.17602/M2/L843966. Specimen photos and a backup of the 3D models are available on Zenodo: https://doi.org/10.5281/zenodo.19596154. Benefits Generated: Benefits from this research accrue from the sharing of data and results on public databases as described above and from the inclusion as co-authors the curator and manager of the museum collection that provided the specimens.

CRediT author contributions

Deon de Jager: Conceptualization; Data curation; Formal analysis; Funding Acquisition; Investigation; Visualization; Writing – Original Draft Preparation. **Andi M. Wilson**: Formal analysis; Writing – Review & Editing; Visualization. **Alba Rey-Iglesia**: Investigation; Writing – Review & Editing. **J. Tyler Faith**: Resources; Writing – Review & Editing. **Kaedan O'Brien**: Resources; Writing – Review & Editing. **Wendy Black**: Resources; Writing – Review & Editing. **Wilhelmina Seconna**: Resources; Writing – Review & Editing. **Olivia Hall**: Investigation; Writing – Review & Editing. **Paul Szpak**: Resources; Writing – Review & Editing. **Eline D. Lorenzen**: Conceptualization; Funding Acquisition;

Resources; Supervision; Writing – Original Draft Preparation.

Funding statement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101026951. This study was supported by the Villum Fonden Young Investigator Programme 37352 and the Carlsberg Foundation grant CF23-1061 to E.D.L.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Heritage Western Cape, the South African Heritage Resources Agency, and the Iziko Museum curators for granting access to the specimens. We also thank the staff at the Geo-Genetics Sequencing Core, Globe Institute, University of Copenhagen for their support, assistance, and service throughout the study. We thank Kasper Lykke Hansen for his assistance with generating 3D scans for selected specimens with the Artec Space Spider, and Theis Zetner Trolle Jensen for use of his camera for Phase 2 specimen photos and training on the 3Shape Trios 3D scanner. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101026951. This study was supported by the Villum Fonden Young Investigator Programme 37352 and the Carlsberg Foundation grant CF23-1061 to E. D.L.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quascirev.2026.110076.

Data availability

A link to the data and/or code is provided as part of this submission.

References

Adler, C.J., Haak, W., Donlon, D., Cooper, A., 2011. Survival and recovery of DNA from ancient teeth and bones. J. Archaeol. Sci. 38 (5), 956–964. https://doi.org/10.1016/j.jas.2010.11.010.

Alberti, F., Gonzalez, J., Paijmans, J.L.A., Basler, N., Preick, M., Henneberger, K., Trinks, A., Rabeder, G., Conard, N.J., Münzel, S.C., Joger, U., Fritsch, G., Hildebrandt, T., Hofreiter, M., Barlow, A., 2018. Optimized DNA sampling of ancient bones using computed tomography scans. Mol. Ecol. Resour. 18 (6), 1196–1208. https://doi.org/10.1111/1755-0998.12911.

Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T.P., Willerslev, E., Zhang, G., Scofield, R.P., Holdaway, R.N., Bunce, M., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc. Biol. Sci. 279 (1748), 4724–4733. https://doi.org/10.1098/rspb.2012.1745.

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., et al., 2015. Population genomics of Bronze Age Eurasia. Nature 522 (7555), 167–172. https://doi.org/10.1038/nature14507.

Bradfield, J., Forssman, T., Spindler, L., Antonites, A.R., 2019. Identifying the animal species used to manufacture bone arrowheads in South Africa. Archaeol. Anthropol. Sci. 11 (6), 2419–2434. https://doi.org/10.1007/s12520-018-0688-5.

Bushnell, B., 2022. *BBMap*. Sourceforge. https://sourceforge.net/projects/bbmap/.

Campos, P.F., Craig, O.E., Turner-Walker, G., Peacock, E., Willerslev, E., Gilbert, M.T.P., 2012. DNA in ancient bone - where is it located and how should we extract it? Anatomischer Anzeiger [Annals of Anatomy] 194 (1), 7–16. https://doi.org/10.1016/j.aanat.2011.07.003.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., Gilbert, M.T.P., 2018. Single-tube library preparation for degraded DNA. Methods Ecol. Evol./British Ecological Society 9 (2), 410–419. https://doi.org/10.1111/2041-210X.12871.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: an open-source mesh processing tool. In: Scarano, V., Chiara, R.D., Erra, U. (Eds.), Eurographics Italian Chapter Conference. The Eurographics Association. https://doi.org/10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136.

Cleghorn, N., Potts, A.J., Cawthra, H.C., 2020. The Palaeo-Agulhas Plain: a lost world and extinct ecosystem. Quat. Sci. Rev. 235, 106308. https://doi.org/10.1016/j.quascirev.2020.106308.

Collins, M.J., Nielsen-Marsh, C.M., Hiller, J., Smith, C.I., Roberts, J.P., Prigodich, R.V., Wess, T.J., Csapò, J., Millard, A.R., Turner-Walker, G., 2002. The survival of organic matter in bone: a review. Archaeometry 44 (3), 383–394. https://doi.org/10.1111/1475-4754.t01-1-00071.

Coutu, A.N., Taurozzi, A.J., Mackie, M., Jensen, T.Z.T., Collins, M.J., Sealy, J., 2021. Palaeoproteomics confirm earliest domesticated sheep in southern Africa ca. 2000 BP. Sci. Rep. 11 (1), 6631. https://doi.org/10.1038/s41598-021-85756-8.

Coutu, A.N., Whitelaw, G., le Roux, P., Sealy, J., 2016. Earliest evidence for the ivory trade in southern Africa: isotopic and ZooMS analysis of seventh–tenth century AD ivory from KwaZulu-Natal. Afr. Archaeol. Rev. 33 (4), 411–435. https://doi.org/10.1007/s10437-016-9232-0.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. 110 (39), 15758–15763. https://doi.org/10.1073/pnas.1314445110.

Dabney, J., Meyer, M., 2019. Extraction of highly degraded DNA from ancient bones and teeth. Methods Mol. Biol. 1963, 25–29. https://doi.org/10.1007/978-1-4939-9176-1_4.

Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., Allentoft, M.E., 2015. Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. 5 (11184), 11184. https://doi.org/10.1038/srep11184.

de Jager, D., Li, V., Cabrera, A.A., Lorenzen, E., 2026. Subsampling Bones and Teeth for Ancient DNA Extraction. Protocols.io. https://doi.org/10.17504/protocols.io.q26g7n5o3lwz/v2.

Deacon, H.J., 1995. Two late Pleistocene-Holocene archaeological depositories from the southern Cape, South Africa. S. Afr. Archaeol. Bull. 50 (162), 121. https://doi.org/10.2307/3889061.

Dobberstein, R.C., Collins, M.J., Craig, O.E., Taylor, G., Penkman, K.E.H., Ritz-Timme, S., 2009. Archaeological collagen: why worry about collagen diagenesis? Archaeol. Anthropol. Sci. 1 (1), 31–42. https://doi.org/10.1007/s12520-009-0002-7.

Eriksen, A.M.H., Rodríguez, J.A., Seersholm, F., Hollund, H.I., Gotfredsen, A.B., Collins, M.J., Grønnow, B., Pedersen, M.W., Gilbert, M.T.P., Matthiesen, H., 2025. Exploring DNA degradation *in situ* and in museum storage through genomics and metagenomics. Commun. Biol. 8 (1), 210. https://doi.org/10.1038/s42003-025-07616-9.

Essel, E., Korlević, P., Meyer, M., 2021. A method for the temperature-controlled extraction of DNA from ancient bones. Biotechniques 71 (1), 382–386. https://doi.org/10.2144/btn-2021-0025.

Essel, E., Zavala, E.I., Schulz-Kornas, E., Kozlikin, M.B., Fewlass, H., Vernot, B., Shunkov, M.V., Derevianko, A.P., Douka, K., Barnes, I., Soulier, M.-C., Schmidt, A., Szymanski, M., Tsanova, T., Sirakov, N., Endarova, E., McPherron, S.P., Hublin, J.-J., Kelso, J., et al., 2023. Ancient human DNA recovered from a Palaeolithic pendant. Nature 618 (7964), 328–332. https://doi.org/10.1038/s41586-023-06035-2.

Faith, J.T., 2013. Taphonomic and paleoecological change in the large mammal sequence from Boomplaas Cave, Western Cape, South Africa. J. Hum. Evol. 65 (6), 715–730. https://doi.org/10.1016/j.jhevol.2013.09.001.

Faith, J.T., Chase, B.M., Pargeter, J., 2024. The Last Glacial Maximum climate at Boomplaas Cave, South Africa. Quat. Sci. Rev. 329 (108557), 108557. https://doi.org/10.1016/j.quascirev.2024.108557.

Feuerborn, T.R., Palkopoulou, E., van der Valk, T., von Seth, J., Munters, A.R., Pečnerová, P., Dehasque, M., Ureña, I., Ersmark, E., Lagerholm, V.K., Krzewińska, M., Rodríguez-Varela, R., Götherström, A., Dalén, L., Díez-del-Molino, D., 2020. Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets. BMC Genom. 21 (1), 844. https://doi.org/10.1186/s12864-020-07229-y.

Fisher, E.C., Bar-Matthews, M., Jerardino, A., Marean, C.W., 2010. Middle and Late Pleistocene paleoscape modeling along the southern coast of South Africa. Quat. Sci. Rev. 29 (11), 1382–1398. https://doi.org/10.1016/j.quascirev.2010.01.015.

Fox, J., Weisberg, S., 2019. An R Companion to Applied Regression (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gansauge, M.-T., Aximu-Petri, A., Nagel, S., Meyer, M., 2020. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nat. Protoc. 15 (8), 2279–2300. https://doi.org/10.1038/s41596-020-0338-0.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., Meyer, M., 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. 45 (10). https://doi.org/10.1093/nar/gkx033 e79–e79.

Gretzinger, J., Gibbon, V.E., Penske, S.E., Sealy, J.C., Rohrlach, A.B., Salazar-García, D. C., Krause, J., Schiffels, S., 2024. 9,000 years of genetic continuity in southernmost Africa demonstrated at Oakhurst rockshelter. Nat. Ecol. Evol. 8 (11), 2121–2134. https://doi.org/10.1038/s41559-024-02532-3.

Götherström, A., Collins, M.J., Angerbjörn, A., Lidén, K., 2002. Bone preservation and DNA amplification. Archaeometry 44 (3), 395–404. https://doi.org/10.1111/1475-4754.00072.

Hempel, E., Bibi, F., Faith, J.T., Koepfli, K.-P., Klittich, A.M., Duchêne, D.A., Brink, J.S., Kalthoff, D.C., Dalén, L., Hofreiter, M., Westbury, M.V., 2022. Blue turns to gray: paleogenomic insights into the evolutionary history and extinction of the blue antelope (*Hippotragus leucophaeus*). Mol. Biol. Evol. 39 (12). https://doi.org/10.1093/molbev/msac241.

Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., Collins, M.J., 2018. A guide to ancient protein studies. Nat. Ecol. Evol. 2 (5), 791–799. https://doi.org/10.1038/s41559-018-0510-x.

Henshilwood, C.S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G.A.T., Mercier, N., Sealy, J.C., Valladas, H., Watts, I., Wintle, A.G., 2002. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. Science (New York, N.Y.) 295 (5558), 1278–1280. https://doi.org/10.1126/science.1067575.

Higham, T., Luftensteiner, K., van der Sluis, L., Giannì, M., Wiegele, M., Papadogianni, A., Steier, P., Gruber, D., Schmidt, B., Schmidt, K., Belinski, A., Mensan, R., Kozlikin, M., Shunkov, M., Neruda, P., Nerudová, Z., Horejs, B., Schulze, J., Douka, K., Stafford Jr., T.W., 2026. Minimally destructive radiocarbon dating of bone. Rapid Commun. Mass Spectrom.: RCM (Rapid Commun. Mass Spectrom.) 40 (9), e70040. https://doi.org/10.1002/rcm.70040.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., Wilson, A.C., 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature 312 (5991), 282–284. https://doi.org/10.1038/312282a0.

Hofreiter, M., Paijmans, J.L.A., Goodchild, H., Speller, C.F., Barlow, A., Fortes, G.G., Thomas, J.A., Ludwig, A., Collins, M.J., 2015. The future of ancient DNA: technical advances and conceptual shifts. Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology 37 (3), 284–293. https://doi.org/10.1002/bies.201400160.

Horsburgh, K.A., 2008. Wild or domesticated? An ancient DNA approach to canid species identification in South Africa's Western Cape Province. J. Archaeol. Sci. 35 (6), 1474–1480. https://doi.org/10.1016/j.jas.2007.10.012.

Horsburgh, K.A., Moreno-Mayar, J.V., Gosling, A.L., 2016. Revisiting the kalahari debate in the highlands: ancient DNA provides new faunal identifications at Sehonghong, Lesotho. Azania 51 (3), 295–306. https://doi.org/10.1080/0067270X.2016.1169041.

Horsburgh, K.A., Gosling, A.L., 2020. Systematic ancient DNA species identification fails to find late Holocene domesticated cattle in southern Africa. Biology 9 (10), 316. https://doi.org/10.3390/biology9100316.

Horsburgh, K.A., Moreno-Mayar, J.V., 2015. Molecular identification of sheep at Blydefontein Rock Shelter, South Africa. South. Afr. Humanit. 27 (1), 65–80. https://journals.co.za/content/nmsa_sah/27/1/EJC179772.

Janzen, A., Richter, K.K., Mwebi, O., Brown, S., Onduso, V., Gatwiri, F., Ndiema, E., Katongo, M., Goldstein, S.T., Douka, K., Boivin, N., 2021. Distinguishing African bovids using zooarchaeology by mass spectrometry (ZooMS): new peptide markers and insights into Iron Age economies in Zambia. PLoS One 16 (5), e0251061. https://doi.org/10.1371/journal.pone.0251061.

Kapp, J.D., Green, R.E., Shapiro, B., 2021. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. J. Hered. 112 (3), 241–249. https://doi.org/10.1093/jhered/esab012.

Klein, R.G., 1972. The late Quaternary Mammalian fauna of Nelson Bay Cave (Cape Province, South Africa): its implications for megafaunal extinctions and environmental and cultural change. Quat. Res. 2 (2), 135–142. https://doi.org/10.1016/0033-5894(72)90034-8.

Klein, R.G., 1974. Environment and subsistence of prehistoric man in the southern Cape Province, South Africa. World Archaeol. 5 (3), 249–284. https://doi.org/10.1080/00438243.1974.9979574.

Klein, R.G., 1984. Mammalian extinctions and Stone Age people in Africa. In: Klein, R.G., Martin, P.S. (Eds.), Quaternary Extinctions. University of Arizona Press, pp. 553–573. https://doi.org/10.2307/j.ctv264f91j.33.

Klein, R.G., Cruz-Uribe, K., 2016. Large mammal and tortoise bones from Elands Bay Cave (South Africa) : implications for Later Stone Age environment and ecology. South. Afr. Humanit. 29 (1), 259–282. https://doi.org/10.10520/EJC-55c1a2524.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., Makowski, D., 2021. Performance: an R package for assessment, comparison and testing of statistical models. J. Open Source Softw. 6 (60), 3139. https://doi.org/10.21105/joss.03139.

Le Meillour, L., Zazzo, A., Zirah, S., Tombret, O., Barriel, V., Arthur, K.W., Arthur, J.W., Cauliez, J., Chaix, L., Curtis, M.C., Gifford-Gonzalez, D., Gunn, I., Gutherz, X., Hildebrand, E., Khalidi, L., Millet, M., Mitchell, P., Studer, J., Vila, E., et al., 2023. The name of the game: palaeoproteomics and radiocarbon dates further refine the presence and dispersal of caprines in eastern and southern Africa. R. Soc. Open Sci. 10 (11), 231002. https://doi.org/10.1098/rsos.231002.

Le Meillour, L., Zirah, S., Zazzo, A., Cersoy, S., Détroit, F., Imalwa, E., Lebon, M., Nankela, A., Tombret, O., Pleurdeau, D., Lesur, J., 2020. Palaeoproteomics gives new insight into early southern African pastoralism. Sci. Rep. 10 (1), 14427. https://doi.org/10.1038/s41598-020-71374-3.

Lebon, M., Reiche, I., Gallet, X., Bellot-Gurlet, L., Zazzo, A., 2016. Rapid quantification of bone collagen content by ATR-FTIR spectroscopy. Radiocarbon 58 (1), 131–145. https://doi.org/10.1017/rdc.2015.11.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25 (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362 (6422), 709–715. https://doi.org/10.1038/362709a0.

Lindahl, A., Epp, L.S., Boessenkool, S., Pedersen, M.W., Brace, S., Heintzman, P.D., Dalén, L., Díez del Molino, D., 2025. Palaeogenomic inference of biodiversity dynamics across Quaternary timescales. Nature Reviews Biodiversity 1 (4), 233–247. https://doi.org/10.1038/s44358-025-00033-0.

Lindahl, T., Andersson, A., 1972. Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. Biochemistry 11 (19), 3618–3623. https://doi.org/10.1021/bi00769a019.

Loftus, E., Sealy, J., Lee-Thorp, J., 2016. New radiocarbon dates and Bayesian models for Nelson Bay Cave and Byneskranskop 1: implications for the South African Later Stone Age sequence. Radiocarbon 58 (2), 365–381. https://doi.org/10.1017/RDC.2016.12.

Mackay, A., Stewart, B.A., Chase, B.M., 2014. Coalescence and fragmentation in the late Pleistocene archaeology of southernmost Africa. J. Hum. Evol. 72, 26–51. https://doi.org/10.1016/j.jhevol.2014.03.003.

Marean, C.W., Cawthra, H.C., Cowling, R.M., Esler, K.J., Fisher, E., Milewski, A., Potts, A. J., Singels, E., De Vynck, J., 2014. Stone Age people in a changing South African Greater Cape Floristic Region. In: Allsopp, Nicky Colville, Verboom, Jonathan F., Anthony, G. (Eds.), Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region. Oxford University Press, pp. 164–199. https://doi.org/10.1093/acprof:oso/9780199679584.003.0008.

Marean, C.W., Cowling, R.M., Franklin, J., 2020. The Palaeo-Agulhas Plain: temporal and spatial variation in an extraordinary extinct ecosystem of the Pleistocene of the cape floristic region. Quat. Sci. Rev. 235, 106161. https://doi.org/10.1016/j.quascirev.2019.106161.

Mathieson, I., Abascal, F., Vinner, L., Skoglund, P., Pomilla, C., Mitchell, P., Arthur, C., Gurdasani, D., Willerslev, E., Sandhu, M.S., Dewar, G., 2020. An ancient baboon genome demonstrates long-term population continuity in Southern Africa. Genome Biol. Evol. 12 (4), 407–412. https://doi.org/10.1093/gbe/evaa019.

McColl, H., Kroonen, G., Moreno-Mayar, J.V., Valeur Seersholm, F., Scorrano, G., Pinotti, T., Vimala, T., Sindbaek, S.M., Ethelberg, P., Fyfe, R., Gaillard, M.-J., Larsen, H.M.E., Mortensen, M.F., Demeter, F., Jorkov, M.L.S., Bergerbrant, S., Damgaard, P. de B., Allentoft, M.E., Vinner, L., et al., 2024. Steppe ancestry in western Eurasia and the spread of the Germanic Languages. bioRxiv. https://doi.org/10.1101/2024.03.13.584607, 2024.03.13.584607.

McGrath, K., Rowsell, K., Gates St-Pierre, C., Tedder, A., Foody, G., Roberts, C., Speller, C., Collins, M., 2019. Identifying archaeological bone via non-destructive ZooMS and the materiality of symbolic expression: examples from Iroquoian bone points. Sci. Rep. 9 (1), 11027. https://doi.org/10.1038/s41598-019-47299-x.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010 (6). https://doi.org/10.1101/pdb.prot5448 db.prot5448.

Miller, G.H., Beaumont, P.B., Deacon, H.J., Brooks, A.S., Hare, P.E., Jull, A.J.T., 1999. Earliest modern humans in Southern Africa dated by isoleucine epimerization in ostrich eggshell. Quat. Sci. Rev. 18 (13), 1537–1548. https://doi.org/10.1016/s0277-3791(99)00044-x.

Neukamm, J., Peltzer, A., Nieselt, K., 2021. DamageProfiler: fast damage pattern calculation for ancient DNA. Bioinformatics 37 (20), 3652–3653. https://doi.org/10.1093/bioinformatics/btab190.

Oliva, A., Tobler, R., Llamas, B., Souilmi, Y., 2021. Additional evaluations show that specific BWA-Aln settings still outperform BWA-Mem for ancient DNA data alignment. Ecol. Evol. 11 (24), 18743–18748. https://doi.org/10.1002/ece3.8297.

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 (335), 108–120. https://doi.org/10.1017/S0003598X00048651.

Oskolkov, N., Jin, C., Clinton, S.L., Guinet, B., Wijnands, F., Johnson, E., Kutschera, V.E., Kinsella, C.M., Heintzman, P.D., van der Valk, T., 2025. Improving taxonomic inference from ancient environmental metagenomes by masking microbial-like regions in reference genomes. GigaScience 14 (giaf108). https://doi.org/10.1093/gigascience/giaf108 giaf108.

Ottoni, C., Koon, H.E.C., Collins, M.J., Penkman, K.E.H., Rickards, O., Craig, O.E., 2009. Preservation of ancient DNA in thermally damaged archaeological bone. Sci. Nat. 96 (2), 267–278. https://doi.org/10.1007/s00114-008-0478-5.

Pálsdóttir, A.H., Bläuer, A., Rannamäe, E., Boessenkool, S., Hallsson, J.H., 2019. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. R. Soc. Open Sci. 6 (10), 191059. https://doi.org/10.1098/rsos.191059.

Poinar, H.N., Höss, M., Bada, J.L., Pääbo, S., 1996. Amino acid racemization and the preservation of ancient DNA. Science 272 (5263), 864–866. https://doi.org/10.1126/science.272.5263.864.

Poinar, H.N., Stankiewicz, B.A., 1999. Protein preservation and DNA retrieval from ancient tissues. Proc. Natl. Acad. Sci. U. S. A 96 (15), 8426–8431. https://doi.org/10.1073/pnas.96.15.8426.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., Geigl, E.-M., 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. Proc. Natl. Acad. Sci. U. S. A 104 (3), 739–744. https://doi.org/10.1073/pnas.0610257104.

R Core Team, 2025. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Radloff, F.G.T., Mucina, L., Bond, W.J., le Roux, P.J., 2010. Strontium isotope analyses of large herbivore habitat use in the Cape Fynbos region of South Africa. Oecologia 164 (2), 567–578. https://doi.org/10.1007/s00442-010-1731-0.

Rector, A.L., Reed, K.E., 2010. Middle and late Pleistocene faunas of Pinnacle Point and their paleoecological implications. J. Hum. Evol. 59 (3), 340–357. https://doi.org/10.1016/j.jhevol.2010.07.002.

Rey-Iglesia, A., de Jager, D., Li, V., Cabrera, A.A., Westbury, M.V., Lorenzen, E.D., 2026. USER II Treatment of Ancient DNA Extracts. Protocols.io. https://doi.org/10.17504/protocols.io.eq2ly6qrwgx9/v2.

Rey-Iglesia, A., de Jager, D., Li, V., Lorenzen, E., 2025. Santa Cruz Reaction (SCR): Single-Stranded Ancient DNA Library Protocol. Protocols.io. https://doi.org/10.17504/protocols.io.5jyl8d4p7g2w/v2.

Reynard, J.P., Wurz, S., 2020. The palaeoecology of Klasies River, South Africa: an analysis of the large mammal remains from the 1984–1995 excavations of Cave 1 and 1A. Quat. Sci. Rev. 237, 106301. https://doi.org/10.1016/j.quascirev.2020.106301.

Richter, K.K., Codlin, M.C., Seabrook, M., Warinner, C., 2022. A primer for ZooMS applications in archaeology. Proc. Natl. Acad. Sci. U. S. A 119 (20), e2109323119. https://doi.org/10.1073/pnas.2109323119.

Rohland, N., Glocke, I., Aximu-Petri, A., Meyer, M., 2018. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. 13 (11), 2447–2461. https://doi.org/10.1038/s41596-018-0050-5.

Rohland, N., Hofreiter, M., 2007. Ancient DNA extraction from bones and teeth. Nat. Protoc. 2 (7), 1756–1762. https://doi.org/10.1038/nprot.2007.247.

Ryder, C., Celis, G., Devièse, T., Talamo, S., Douka, K., Collins, M., Perri, A., Thakar, H., Pestle, W., Sponheimer, M., 2026. Refining near-infrared spectroscopy for collagen quantification: a new predictive model for archaeological bone. J. Archaeol. Sci. 185 (106448), 106448. https://doi.org/10.1016/j.jas.2025.106448.

Scarsbrook, L., Verry, A.J.F., Walton, K., Hitchmough, R.A., Rawlence, N.J., 2023. Ancient mitochondrial genomes recovered from small vertebrate bones through minimally destructive DNA extraction: phylogeography of the New Zealand gecko genus *Hoplodactylus*. Mol. Ecol. 32 (11), 2964–2984. https://doi.org/10.1111/mec.16434.

Schubert, M., Ermini, L., Sarkissian, C.D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., Orlando, L., 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. 9 (5), 1056–1082. https://doi.org/10.1038/nprot.2014.063.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A.S., Willerslev, E., Krogh, A., Orlando, L., 2012. Improving ancient DNA read mapping against modern reference genomes. BMC Genom. 13 (1), 178. https://doi.org/10.1186/1471-2164-13-178.

Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9 (1), 88. https://doi.org/10.1186/s13104-016-1900-2.

Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A.D., Bada, J., Poinar, H., 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res. 37 (10), 3215–3229. https://doi.org/10.1093/nar/gkp159.

Schweitzer, F.R., 1979. Excavations at Die Kelders, Cape Province, South Africa. The Holocene deposits. Ann. S. Afr. Mus. 78, 101–233. *Annale van Die Suid-Afrikaanse Museum*. https://www.biodiversitylibrary.org/part/74632.

Sealy, J., Naidoo, N., Hare, V.J., Brunton, S., Faith, J.T., 2020. Climate and ecology of the Palaeo-Agulhas Plain from stable carbon and oxygen isotopes in bovid tooth enamel from Nelson Bay Cave, South Africa. Quat. Sci. Rev. 235, 105974. https://doi.org/10.1016/j.quascirev.2019.105974.

Sillen, A., Parkington, J., 1996. Diagenesis of bones from Eland's Bay Cave. J. Archaeol. Sci. 23 (4), 535–542. https://doi.org/10.1006/jasc.1996.0050.

Singer, R., Wymer, J., 1982. The Middle Stone Age at Klasies River Mouth in South Africa. University of Chicago Press. https://books.google.dk/books?id=AO8eah2dLFgC.

Sosa, C., Vispe, E., Núñez, C., Baeta, M., Casalod, Y., Bolea, M., Hedges, R.E.M., Martinez-Jarreta, B., 2013. Association between ancient bone preservation and dna yield: a multidisciplinary approach: Bone preservation and DNA yield. Am. J. Phys. Anthropol. 151 (1), 102–109. https://doi.org/10.1002/ajpa.22262.

van Doorn, N.L., Hollund, H., Collins, M.J., 2011. A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. Archaeol. Anthropol. Sci. 3 (3), 281–289. https://doi.org/10.1007/s12520-011-0067-y.

Vogel, J., 2001. Radiometric dates for the Middle Stone Age in South Africa. In: Tobias, Phillip V. Raath, Moggi-Cecchi, Michael A., Doyle, Jacopo, Gerald, A. (Eds.), Humanity from African Naissance to Coming Millennia : Colloquia in Human Biology and Palaeoanthropology. Firenze University Press ; Witwatersrand University Press, pp. 261–268. https://doi.org/10.1400/38989 *(Monografie. Scienze ; 1.*

Wadsworth, C., Buckley, M., 2018. Characterization of proteomes extracted through collagen-based stable isotope and radiocarbon dating methods. J. Proteome Res. 17 (1), 429–439. https://doi.org/10.1021/acs.jproteome.7b00624.

Wales, N., Carøe, C., Sandoval-Velasco, M., Gamba, C., Barnett, R., Samaniego, J.A., Madrigal, J.R., Orlando, L., Gilbert, M.T.P., 2015. New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. Biotechniques 59 (6), 368–371. https://doi.org/10.2144/000114364.

Westbury, M.V., Rey-Iglesia, A., Cabrera, A.A., Li, V., de Jager, D., Lorenzen, E.D., 2025. Ancient DNA extraction from bones and teeth. Protocols.io. https://doi.org/10.17504/protocols.io.x54v9ronpv3e/v1.

Westbury, M.V., Rey-Iglesia, A., Li, V., de Jager, D., Lorenzen, E.D., 2026. Blunt-end-single-tube (BEST): double-stranded ancient DNA library protocol. Protocols.io. https://doi.org/10.17504/protocols.io.4r3l21yq4g1y/v1.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., et al., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (Issue 43), 1686. https://doi.org/10.21105/joss.01686.

Wood, R., 2015. From revolution to convention: the past, present and future of radiocarbon dating. J. Archaeol. Sci. 56, 61–72. https://doi.org/10.1016/j.jas.2015.02.019.