

https://dspace.org/

Colonialdriven extinction of the blue antelope despite genomic adaptation to low populationsize

E Hempel; John Faith; M Preick; M Westbury; D de Jager ...

CELL PRESS
https://hdl.handle.net/10539/41818
WOS

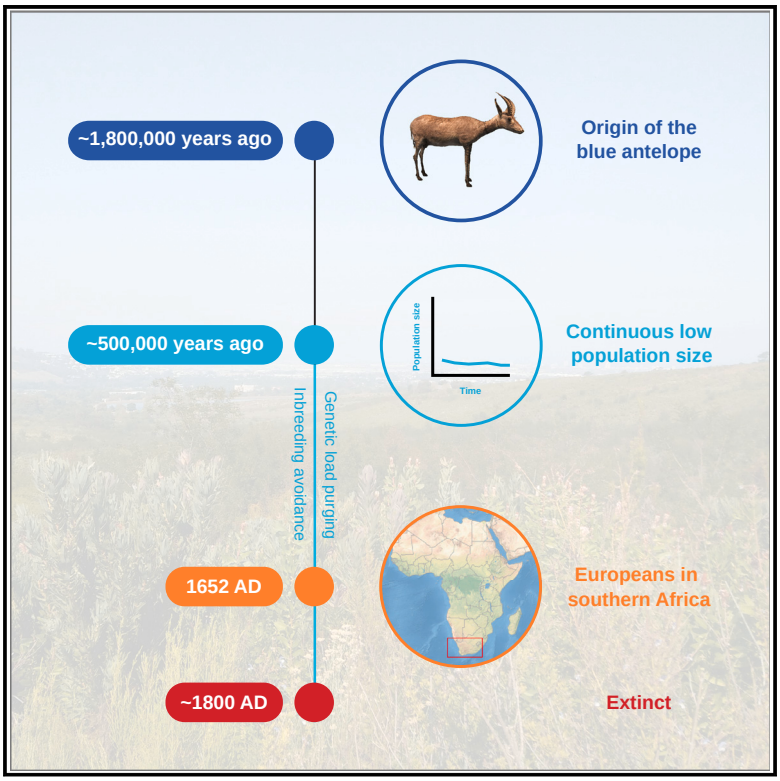
Downloaded from a DSpace repository. https://dspace.org/

Report

Current Biology

Colonial-driven extinction of the blue antelope despite genomic adaptation to low population size

Graphical abstract



Highlights

- Blue antelope had low diversity and population size for at least 400,000 years
- Lack of inbreeding/high levels of purging suggest adaptation to low population size
- Long-term effective population size unaffected by glacial cycles/habitat size
- Demonstrates the impact of European colonization on African wildlife

Authors

Elisabeth Hempel, J. Tyler Faith, Michaela Preick, ..., Love Dalén, Michael V. Westbury, Michael Hofreiter

Correspondence

hempel.elisabeth@posteo.org (E.H.), love.dalen@zoologi.su.se (L.D.), m.westbury@sund.ku.dk (M.V.W.), michael.hofreiter@uni-potsdam.de (M.H.)

In brief

Hempel et al. show that prior to its extinction, the blue antelope had a low population size and diversity over millennia, unaffected by glacial cycles. However, no signs of inbreeding and extensive purging indicate adaptation to low genomic diversity. Hence, it likely became extinct due to a fast-acting disturbance in the form of European colonists.

Hempel et al., 2024, Current Biology *34*, 2020–2029
May 6, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2024.03.051

CellPress

CellPress
OPEN ACCESS

Current Biology

Report

Colonial-driven extinction of the blue antelope despite genomic adaptation to low population size

Elisabeth Hempel,[1,2,*] J. Tyler Faith,[3,4,5] Michaela Preick,[1] Deon de Jager,[6] Scott Barish,[7] Stefanie Hartmann,[1] José H. Grau,[8,9] Yoshan Moodley,[10] Gregory Gedman,[11] Kathleen Morrill Pirovich,[7,11] Faysal Bibi,[2] Daniela C. Kalthoff,[12] Sven Bocklandt,[7] Ben Lamm,[7] Love Dalén,[13,14,15,*] Michael V. Westbury,[6,16,*] and Michael Hofreiter[1,16,17,*]

[1]Evolutionary Adaptive Genomics, Institute of Biochemistry and Biology, Faculty of Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
[2]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstraße 43, 10115 Berlin, Germany
[3]Natural History Museum of Utah, University of Utah, 301 Wakara Way, Salt Lake City, UT 84108, USA
[4]Department of Anthropology, University of Utah, 260 South Central Campus Drive, Salt Lake City, UT 84112, USA
[5]Origins Centre, University of the Witwatersrand, 2000 Johannesburg, Republic of South Africa
[6]Globe Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark
[7]Colossal Biosciences, Dallas, TX 75247, USA
[8]Center for Species Survival, Smithsonian Conservation Biology Institute, Washington, DC 20008, USA
[9]Amedes Genetics, Amedes Medizinische Dienstleistungen GmbH, 10117 Berlin, Germany
[10]Department of Biological Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa
[11]Form Bio, Dallas, TX 75247, USA
[12]Swedish Museum of Natural History, Department of Zoology, Box 50007, 10405 Stockholm, Sweden
[13]Swedish Museum of Natural History, Department of Bioinformatics and Genetics, Box 50007, 10405 Stockholm, Sweden
[14]Centre for Palaeogenetics, Svante Arrhenius väg 20c, 10691 Stockholm, Sweden
[15]Department of Zoology, Stockholm University, 10691 Stockholm, Sweden
[16]Senior author
[17]Lead contact
*Correspondence: hempel.elisabeth@posteo.org (E.H.), love.dalen@zoologi.su.se (L.D.), m.westbury@sund.ku.dk (M.V.W.), michael.hofreiter@uni-potsdam.de (M.H.)
https://doi.org/10.1016/j.cub.2024.03.051

SUMMARY

Low genomic diversity is generally indicative of small population size and is considered detrimental by decreasing long-term adaptability.[1–6] Moreover, small population size may promote gene flow with congeners and outbreeding depression.[7–13] Here, we examine the connection between habitat availability, effective population size (N_e), and extinction by generating a 40× nuclear genome from the extinct blue antelope (*Hippotragus leucophaeus*). Historically endemic to the relatively small Cape Floristic Region in southernmost Africa,[14,15] populations were thought to have expanded and contracted across glacial-interglacial cycles, tracking suitable habitat.[16–18] However, we found long-term low N_e, unaffected by glacial cycles, suggesting persistence with low genomic diversity for many millennia prior to extinction in ∼AD 1800. A lack of inbreeding, alongside high levels of genetic purging, suggests adaptation to this long-term low N_e and that human impacts during the colonial era (e.g., hunting and landscape transformation), rather than longer-term ecological processes, were central to its extinction. Phylogenomic analyses uncovered gene flow between roan (*H. equinus*) and blue antelope, as well as between roan and sable antelope (*H. niger*), approximately at the time of divergence of blue and sable antelope (∼1.9 Ma). Finally, we identified the *LYST* and *ASIP* genes as candidates for the eponymous bluish pelt color of the blue antelope. Our results revise numerous aspects of our understanding of the interplay between genomic diversity and evolutionary history and provide the resources for uncovering the genetic basis of this extinct species' unique traits.

RESULTS AND DISCUSSION

Genomic diversity is considered a key component of biodiversity[19] and is widely used as a proxy for population size in evolutionary and conservation genetics research.[1] Low population size is regarded as problematic as it is expected to facilitate inbreeding, a reduction in fitness and adaptability, and an accumulation of deleterious mutations.[2–6] Moreover, low population size is assumed to increase the probability of gene flow between related species (hybridization) due to the lack of available conspecific mating partners (Hubbs principle[7]).[8–10] This, however, might lead to outbreeding depression and loss of locally adapted variation.[11–13] Although various factors can influence population size, the availability of suitable habitat is generally considered to be of key importance. Therefore, population size is expected to change according to habitat availability.[20]

 © 2024 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

CellPress
OPEN ACCESS

The blue antelope, *Hippotragus leucophaeus* (Pallas, 1766), is an ideal species to investigate the relationships between population size and genetic load, congeneric gene flow, and habitat availability. This species was endemic to the Cape Floristic Region (CFR) in southernmost Africa[14,15,21] and is so far the only large African mammal species documented to have gone extinct since 1500.[22,23] The paleoecological history of the CFR raises the possibility that the blue antelope's recent extinction was, in part, a consequence of longer-term ecosystem changes. Previous work in the CFR has emphasized the abundance and diversity of grazers, such as black wildebeest (*Connochaetes gnou*), buffalo (*Syncerus caffer* and *S. antiquus*), and equids (*Equus quagga*, *E. zebra*, and *E. capensis*) in late Pleistocene fossil assemblages, distinct from the relatively species-poor, browser-dominated faunas through the Holocene.[24,25] It is thought that the grazer-dominated faunas of the Pleistocene relate to the expansion of grassy habitats during glacial phases paired with the expanded landmass provided by the ~85,000 km^2 Palaeo-Agulhas Plain that was exposed during marine regressions.[16–18,24,26] Fossil evidence demonstrates that the blue antelope, like other grazers, was more abundant and broadly distributed during the Pleistocene.[21] Due to its presumed ecological link to relatively grassy habitats, the blue antelope's population size is expected to have tracked glacial-interglacial cycles, with a higher population size during glacial phases. A previous study showed that the blue antelope had low genetic diversity in the early Holocene around 9,500 years ago, which suggests it already had a low population size some time before then.[27] However, due to limited data quantity, the previous study was not able to investigate its population size dynamics or the potential impact of low diversity on the species.

Here, we investigated the evolution and eventual extinction of the blue antelope at greater depth by generating a high-coverage blue antelope genome from a historical specimen. As we used the specimen from a previous study, which was shown to have unusual guanine-to-thymine damage patterns, we first investigated which mapping approach would yield the best result by simulating data showing the damage patterns found in our individual (Figure S1).[27] From the seven approaches tried, we found bwa aln −o 2 −n 0.01 to result in the highest number of mapped base pairs (i.e., BWA1, Table S2). Thus, this mapping approach minimizes the reference biases, which can cause alternative alleles and therefore also reads with DNA damage-induced substitutions to not map as efficiently as the reference allele.[28] Using this mapping approach and the scimitar-horned oryx as the reference genome, we successfully mapped 2,356,622,046 unique reads out of the total of 23,454,910,916 raw sequencing reads, corresponding to an endogenous content of ~10% and a 40.12× coverage genome. The average read length for mapped reads was 46 bp.

During the mapping test, we performed genotype calling and noticed a high level of false positive heterozygous base calls in our simulated haploid data, likely caused by our mapping approach, which also allows reads with DNA damage-induced substitutions to be mapped to the reference genome. These false positive heterozygous base calls could, however, be removed when using a minor allele frequency threshold of 0.3. Further investigations into our blue antelope data revealed that both damage and sequencing errors were also inflating heterozygous base calls when the minor allele frequency threshold was set too low (<0.3) (Figure S2). Therefore, we applied a minor allele frequency of >0.3 based on directly counting all possible nucleotides on our empirical data. Using this approach, we found heterozygosity was very low in the blue antelope relative to the roan (*Hippotragus equinus*) and the sable antelope (*Hippotragus niger*), with the roan exhibiting the highest value (Table 1).

Demographic history analyses using the pairwise sequential Markovian coalescence (PSMC) model revealed that the low heterozygosity in the blue antelope was likely a long-term trend coupled with long-term low effective population size (N_e) (Figure 1). Despite the low genome-wide heterozygosity in our blue antelope individual and the low and continuously decreasing N_e for several hundreds of thousands of years, we found no signs of inbreeding when investigating the frequency and length of runs of homozygosity (ROHs) (Table 1; Figures 1 and S3). This lack of ROH, despite low genome-wide levels of heterozygosity, is somewhat unexpected. However, it could be explained by behavioral adaptations preventing mating between kin, such as dispersal from the natal group, which occurs in many antelope species. In fact, the blue antelope's closest living relatives, the sable and roan antelope, both drive young males and sometimes also young females from their natal female groups.[29,30] Furthermore, our investigation of the genetic load of the blue antelope revealed relatively fewer loss-of-function variants than in the roan or the sable antelope, both for the heterozygous and for the homozygous state. This implies purging of both masked and realized genetic load (Table 1) as had already been hypothesized but not tested for this species.[27] High levels of purging of deleterious mutations can be facilitated by long-term low population size, as deleterious mutations are more often present in a homozygous state and consequently more readily selected against than in larger populations. Therefore, from a genomic point of view, if a population manages to survive at low population numbers for a long time, specific adaptations may dampen the effects of genomic erosion, meaning low population size becomes less disadvantageous for the survivability of a population over time. In the absence of extremely fast-acting threats like human impacts or environmental catastrophes, there is increasing evidence that species can indeed survive for hundreds of thousands of years at low population size.[31–36] One well-known example is the Channel Island fox (*Urocyon littoralis*), which is hypothesized to have survived with long-term low population size and diversity due to high levels of purging of deleterious recessive alleles. This is thought to be aiding in its current demographic recovery.[37,38]

Our demographic history reconstruction shows that the blue antelope's N_e does not track Pleistocene glacial cycles (Figure 1); instead, we see a steady decline in N_e since at least 400,000 years ago. The minimum N_e was reached around 20,000 years ago, i.e., during the Last Glacial Maximum, but one should not overinterpret the slight increase in N_e seen after this point in time since PSMC reconstructions become unreliable close to the present.[40] In fact, the fossil record of the blue antelope shows that it became progressively less abundant after the Pleistocene-Holocene transition.[41,42] The observation that the N_e of the blue antelope was low and slowly decreasing for around 400,000 years is inconsistent with the expectation that

Table 1. Heterozygosity and genetic load

		Blue antelope	Sable antelope	Roan antelope
Total # of sites	–	2,048,015,341	2,332,760,508	2,350,448,032
# Heterozygous sites	–	800,433	3,855,115	4,578,111
Heterozygosity	–	0.00039	0.00165	0.00195
# Synonymous-derived homozygous SNPs	–	118,427	171,475	179,171
Realized LoF	–	0.0138	0.0171	0.0171
Masked LoF	–	0.0024	0.0059	0.0062
ROH > 100 kb	total kbp	7,700	155,500	379,000
	% of genome	0.32	6.37	15.54
ROH > 1 Mb	total kbp	0	25,300	47,900
	% of genome	0	1.04	1.96

Estimated autosomal heterozygosity, runs of homozygosity (ROHs), and genetic load for the blue, roan, and sable antelope. Realized LoF was calculated by dividing the number of homozygous-derived LoF SNPs by the total number of synonymous SNPs. Masked LoF was calculated by dividing the total number of heterozygous LoF SNPs by the total number of synonymous SNPs. LoF, loss of function.

population size would have varied in response to glacial-interglacial cycles. Fluctuations in exposed landmass and vegetation due to glacial-interglacial cycles seem to have had no major influence on the blue antelope's N_e. Though our observations are seemingly inconsistent with observed changes in mammal community composition through time in the CFR, whereby grazer abundance and diversity increase during glacial phases and decline during interglacials, they add to recent evidence suggesting that current models of long-term ecosystem dynamics in the region may be in need of refinement. For example, contrary to the expectation that glacial phases supported widespread grassy vegetation, palynological evidence from a ~300,000-year-old marine core ~160 km southwest of Cape Town indicates that grasses in the CFR were consistently rare and that their abundance did not respond to glacial-interglacial cycles.[43] This apparent lack of grassy vegetation is difficult to reconcile with the abundance and diversity of presumed grazers in many Pleistocene fossil assemblages (e.g., Klein,[24] Faith,[25] and Klein et al.[44]), including the rich fossil record of the blue antelope from the southwestern Cape.[24,44] However, there is some evidence that Pleistocene species thought to be grazers were consuming more dicots than expected in the CFR, implying that the extent and importance of grassy vegetation may have been overstated.[45,46] With this in mind, it is possible that the apparent decline in the abundance and distribution of blue antelope fossils across the end of the Pleistocene through the Holocene[24,41] is not a consequence of ecological dynamics linked to 100,000-year glacial-interglacial cycles but perhaps a longer-term process playing out since the middle Pleistocene. Revealing the causes behind the blue antelope's Pleistocene decline will require a better understanding of the paleoecology of the blue antelope, as well as of the ecological dynamics of the CFR.

However, our results allow conclusions about the potential drivers of the blue antelope's extinction. As outlined above, it appears unlikely that ecological changes across the Pleistocene-Holocene transition, proposed by some to have set the stage for its extinction later in the Holocene,[47] played a key role. Likewise, blue antelope and hominins have overlapped for the past ~1 Ma,[44] with late Pleistocene and Holocene archaeological sites from the CFR documenting a long history of human predation on blue antelope (e.g., Klein[21,24,48] and Faith[49]). Our analyses provide no evidence that increased predator-prey interactions in the past ~125,000 years, as suggested by the archaeological record, was associated with a marked decline in population sizes, implying that the blue antelope was able to tolerate human predators armed with a Stone Age toolkit. In addition, the declining and persistently low population sizes over the past 400,000 years further suggest that small populations are not a consequence of the introduction of livestock to the CFR ~2,000 years ago. It is possible that competition with livestock may have further reduced blue antelope populations, rendering them more vulnerable to extinction later on,[42,48] but testing this would require analyses of additional fossil samples within the last 2,000 years. What remains unambiguous, however, is that European colonization of southern Africa in AD 1652 was followed by both landscape transformation and hunting with firearms. The extinction of a species that endured for millenia at low population sizes only ~150 years into the colonial era strongly suggests that these recent human impacts were the decisive factor. A similar fate was met by the quagga (*Equus quagga quagga*), a subspecies of the plains zebra, which was hunted to extinction at the end of the 19th century.[50] Other ungulate grazers also endemic to the CFR, including the bontebok (*Damaliscus pygargus pygargus*) and the Cape mountain zebra (*Equus zebra zebra*), nearly followed the blue antelope and quagga to extinction, and likely for the same reason. Bontebok populations were reduced to fewer than 100 individuals in the early 1800s,[51] and Cape mountain zebra populations were similarly reduced by the 1950s.[52] Both only avoided extinction thanks to dedicated conservation efforts, including the establishment of Bontebok National Park and Mountain Zebra National Park (e.g., van Rensburg[51] and Parker[53]).

Theory suggests that gene flow from a close relative into a species is most likely when the population size of the species in question is low (Hubbs principle[7]).[8–10] Therefore, the previously observed signal of gene flow from the roan into the blue antelope[27] may have been a continuous phenomenon due to the continuously low population size of the blue antelope (Figure 1). Utilizing our new genome and sliding window phylogenomics,

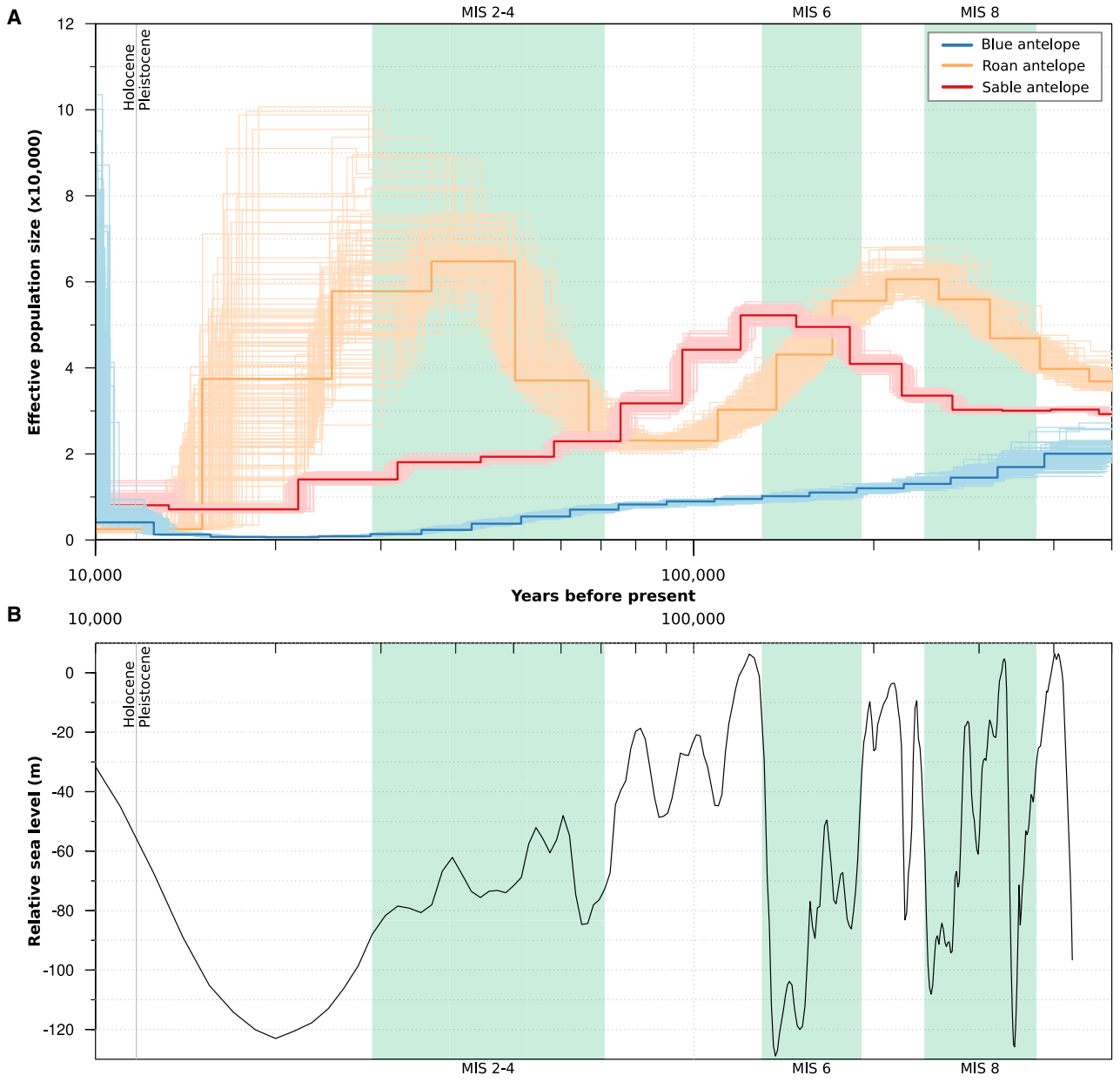


Figure 1. Pairwise sequential Markovian coalescent model for the blue, roan, and sable antelope and relative sea levels

(A) Effective population size (N_e) of the blue, roan, and sable antelope with 100 partial subsamples (generation time blue and sable: 7.1 years, roan antelope: 7.5 years; mutation rate per generation [μ] blue: 1.14×10^{-8}, sable: 1.19×10^{-8}, roan: 1.26×10^{-8}). (B) Average relative sea levels over time.[39] Selected marine isotope stages (MIS) during which the Palaeo-Agulhas Plain should have been available as habitat are marked in light green.

we find support for the inferred species tree (blue and sable antelope as sister species[27,54]) with 52% of gene trees supporting this topology. By contrast, 31% support the blue and the roan antelope as sister species, and 17% support a sister-group relationship between the roan and the sable antelope (Figure 2). We note that due to mapping to the outgroup and the high levels of damage in our blue antelope specimen, there may be a slight bias toward the reference allele in the blue antelope, which could have erroneously inflated the number of windows supporting a closer relationship between the blue antelope and the scimitar-horned oryx and therefore also the number of windows supporting a sister-group relationship between the roan and sable antelope as a result.

The alternative topologies may be driven by incomplete lineage sorting (ILS) and/or gene flow. However, the differences in frequency of the two alternative topologies (31% vs. 17%) suggest at least some post-speciation gene flow between the roan and the blue antelope. Therefore, we evaluated whether the discordant topologies originated from ILS alone or from ILS and gene flow using the quantification of introgression via branch lengths (QuIBL). The results suggest that there was gene flow between the roan and the blue antelope, as well as between the roan and

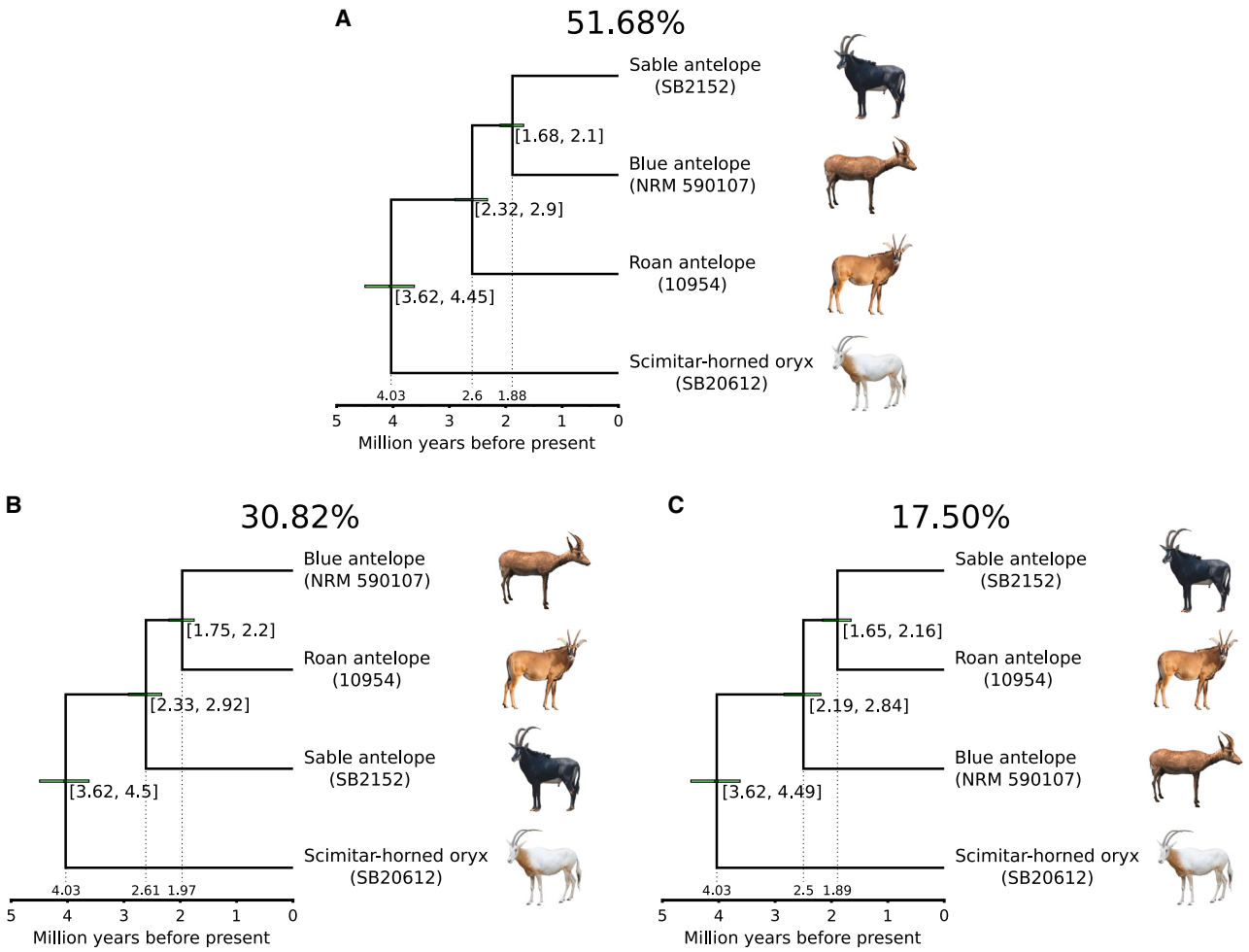


Figure 2. Fossil-calibrated phylogenies from the sliding window tree analysis with frequencies of occurrence

(A) Species phylogeny. (B) Alternative topology with the blue and roan antelope as sister taxa. (C) Alternative topology with the sable and roan antelope as sister taxa. The percentage of gene trees supporting each topology is shown above each tree. Numbers next to nodes indicate the 95% credibility intervals of the divergence time in million years. Numbers above the timescale represent the mean of the respective divergence time of that node. The maximum likelihood phylogenies were fossil calibrated using a uniform prior between 3.6 and 4.5 Ma for the *Oryx*/*Hippotragus* split with a soft maximum bound[55–57]; for further explanations, see Bibi.[58] Photo credits: NRM 590107: Hempel et al.[54]; roan antelope: Charles J. Sharp, Wikimedia Commons, CC-BY 4.0[27]; sable antelope: Paulmaz, Wikimedia Commons, CC-BY 3.0[27]; scimitar-horned oryx: E. Hempel.[27]

the sable antelope (Table S3). We confirmed the finding of gene flow between the blue and roan antelope using *D* statistics. Utilizing the input topology ((blue, sable), roan), we retrieved a *D* value of −0.071 and *Z* score of −23.90. The negative *Z* score and $|Z| > 3$ indicate gene flow between the blue and roan antelope. However, although we know there was gene flow, from these analyses we cannot make inferences about the timing of the gene flow.

We subsequently attempted relative dating of gene flow events using two different approaches: utilizing (1) the length of SNP tracts, and (2) fossil-calibrated phylogenies for each of the three given topologies. When analyzing SNP tracts, we found 32,303 tracts placing the blue and roan antelope as sister species (Figure 2B) and 15,300 tracts placing the sable and roan antelope as sister species (Figure 2C). The median tract lengths for both of these topologies were ∼1,000 bp (Table 2). These alternative topology tracts covered 41,741,868 and 18,469,280 bp of the genome. We also found 7,844 SNP tracts supporting the species tree topology with a median length of 5,064 bp, covering a total of 223,819,284 bp of the genome. The similarity in length of the alternative topology tracts suggests gene flow between the blue and the roan antelope as well as between the sable and the roan antelope to have occurred at approximately the same time, assuming a similar rate of linkage disequilibrium decay across the three species. In accordance with this conclusion, the fossil-calibrated phylogenies for each of the three topologies resulted in similar ages of the most recent common ancestor of the respective sister lineages of around 1.9 million years ago, including very similar and largely overlapping confidence intervals (Figure 2). Similar to the SNP tracts, this suggests that gene flow between the roan antelope and the other two species occurred around the time of divergence between the sable and the blue antelope. Although our PSMC analyses do not extend to the time of divergence of the three lineages and we thus have no estimate for their population size at this time, a

Table 2. SNP tract analysis

	(((blue, sable), roan), outgroup)	(((blue, roan), sable), outgroup)	(((sable, roan), blue) outgroup)
Median tract length	7,844	1,064	1,004
# Tracts	5,064	32,303	15,300
bp covered	223,819,284	41,741,868	18,469,280
Total scored sites	1,628,286	1,212,829	915,877
Minimum tract length	734	70	15
1st quantile tract length	16,375	715	658
Mean tract length	44,202	1,292	1,207
3rd quantile tract length	50,656	1,611	1,517
Maximum tract length	756,638	10,444	23,139

Results of the SNP tract analysis.

possible explanation for the stronger signal of gene flow in the blue antelope compared with the sable antelope is that the ancestral blue antelope population was already small and thus gene flow from the roan antelope had a larger effect than it had on the possibly larger sable antelope population. In the absence of molecular dating approaches, our sliding window phylogenomic and significant *D* statistics results could be interpreted as gene flow via secondary contact long after initial divergence. However, since the gene flow we detected dates to approximately the divergence of the blue and sable antelope, it indicates that gene flow with the roan antelope has likely not occurred since the initial divergence of blue and sable antelope. This result suggests caution should be taken when using secondary contact substantially after divergence as the default interpretation for analyses that do not take the timing of gene flow into account.

The pelt color of the blue antelope is assumed to have expressed a bluish/gray coloration,[23,59] whereas museum specimens now appear as a dull brown color, perhaps due to environmental degradation of pigment in those pelts. Blue or gray coat color in mammals, also known as eumelanin dilution, can be caused by a variety of genes, including *RAB27A*, *MYO5A*, and *MLPH* that affect melanosome transport to keratinocytes,[60,61] downstream genes like *MREG* and *TPCN2*,[62,63] as well as upstream ones like *PMEL* and *LYST* that are linked to melanosome biogenesis and lysosomal trafficking.[64,65] Variants in these genes have been associated with eumelanin dilution in a wide array of species, predominantly domesticates.[60–63,66,67] Key genes involved in melanogenesis and pigmentation patterns in mammals—like *MITF*, *ASIP*, *MC1R*, *TYR*, and *TYRP1*[68]—could also be responsible for the blue antelope's titular coloration by influencing where, which, and how much pigment is produced. When we investigated these candidate genes in the blue antelope genome, we found 24 homozygous, non-synonymous substitutions compared with the roan antelope that were predicted as deleterious by SIFT[69] in 14 genes (*ASIP*, *GPR143*, *BCL2*, *HPS1*, *HPS3*, *HPS4*, *HPS5*, *HPS6*, *LYST*, *AP3D1*, FIG4, *SLC45A2*, *BLOC1S4*, and *ATP7A*) (Table S4). Four of these homozygous deleterious missense variants occur in *LYST* (*lysosomal trafficking regulator*, p.Pro423Thr, p.Val1079Met, p.Ser2827Thr, p.Cys1352Tyr). These protein-coding variants are clear candidates for the blue antelope's bluish hue, although further experimental verification of their phenotypic impact will be necessary. We also found a missense variant close to the start of the *ASIP* protein (p.Arg10Trp) strongly predicted to disrupt its function by SIFT.[69] Loss of function in *ASIP* may contribute to shifting the balance of melanin production from the reddish pheomelanin toward exclusively black eumelanin expression,[70] which contrasts with the roan antelope that predominantly expresses pheomelanin and the sable antelope, which may express full-body eumelanin pigmentation, but only in a sex- and age-dependent manner.[10] However, a deleterious missense mutation that early in the *ASIP* protein (residue 10) points to loss of function of *ASIP*, which would make eumelanin expression invariant to regulation and result in constitutive expression.[68,71] Although we cannot exclude the influence of further, as yet undetected variants in other genes and possibly regulatory regions, the detected combination of sequence variants in *LYST* and *ASIP* could well explain the bluish coat color phenotype that was conspicuous enough to become immortalized in the blue antelope's name even beyond its physical extinction.

In conclusion, our study shows that the blue antelope had low and declining effective population size and genomic diversity for at least 400,000 years prior to its extinction around AD 1800. Contrary to expectations for a large mammalian herbivore, its N_e appeared unaffected by the environmental fluctuations that accompanied glacial-interglacial cycles. Our finding of no signs of inbreeding but high levels of purging suggests that the blue antelope was adapted to its low diversity, showing that, in the absence of fast-acting disturbances, species can survive for many millennia at low population size. Finally, our sliding window fossil-calibrated phylogenies suggest that gene flow analyses that do not consider the timing of gene flow should not be interpreted as evidence for introgression substantially after divergence of the lineages investigated. Overall, our study demonstrates the massive effect of European colonization on southern African wildlife while also providing new insights into the interactions between genomic diversity and evolutionary history that help reveal the unique traits of this extinct species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Lab procedures
  - Bioinformatic procedures
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Estimation of effective population size
  - Heterozygosity
  - Runs of homozygosity
  - Genetic load
  - Fossil calibrated phylogenies
  - Gene flow
  - Coat color analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.03.051.

ACKNOWLEDGMENTS

Part of the computations were done on the HPC Cluster of the University of Potsdam, and we gratefully acknowledge the ZIM team for this support. Computations were also performed on the Mjolnir computing cluster at the University of Copenhagen, and we would like to thank Bent Petersen for continued effort in supporting work on the cluster. The authors also thank David A. Duchêne for support in building the fossil-calibrated phylogenies. The authors acknowledge support from the Science for Life Laboratory, the National Genomics Infrastructure, and UPPMAX for providing assistance with massive parallel sequencing and computational infrastructure.

AUTHOR CONTRIBUTIONS

Conceptualization, E.H., S. Bocklandt, B.L., M.V.W., and M.H.; methodology, E.H., S.H., G.G., K.M.P., M.V.W., and M.H.; software, S.H., J.H.G., and M.V.W.; formal analysis, E.H., D.d.J., S. Barish., and M.V.W.; investigation, E.H., M.P., D.d.J., S. Barish, and M.V.W.; resources, B.L. and M.H.; data curation, E.H. and M.V.W.; writing – original draft, E.H., J.T.F., M.V.W., and M.H.; writing – review & editing, E.H., J.T.F., D.d.J., S. Barish, S.H., Y.M., K.M.P., G.G., F.B., D.C.K., S. Bocklandt, M.V.W., and M.H.; visualization, E.H. and M.V.W.; supervision, M.V.W. and M.H.; project administration, S. Bocklandt, B.L., M.V.W., and M.H.; funding acquisition B.L., L.D. and M.H.

DECLARATION OF INTERESTS

M.H. and L.D. are on the advisory board of Colossal Biosciences and hold stock options. S. Bocklandt, K.M.P., G.G., and S. Barish are employed by Colossal Biosciences or FormBio. B.L. is CEO of Colossal Biosciences.

Received: December 20, 2023
Revised: February 9, 2024
Accepted: March 25, 2024
Published: April 12, 2024

REFERENCES

1. Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. Conserv. Biol. *10*, 1500–1508. https://doi.org/10.1046/j.1523-1739.1996.10061500.x.
2. Caughley, G. (1994). Directions in conservation biology. J. Anim. Ecol. *63*, 215–244. https://doi.org/10.2307/5542.
3. Lynch, M., Conery, J., and Burger, R. (1995). Mutation accumulation and the extinction of small populations. Am. Nat. *146*, 489–518. https://doi.org/10.1086/285812.
4. Lacy, R.C. (1997). Importance of genetic variation to the viability of mammalian populations. J. Mammal. *78*, 320–335. https://doi.org/10.2307/1382885.
5. Reed, D.H., and Frankham, R. (2003). Correlation between fitness and genetic diversity. Conserv. Biol. *17*, 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x.
6. Frankham, R. (2005). Genetics and extinction. Biol. Conserv. *126*, 131–140. https://doi.org/10.1016/j.biocon.2005.05.002.
7. Hubbs, C.L. (1955). Hybridization between fish species in nature. Syst. Biol. *4*, 1–20. https://doi.org/10.2307/sysbio/4.1.1.
8. Jansson, G., Thulin, C.-G., and Pehrson, Å. (2007). Factors related to the occurrence of hybrids between brown hares *Lepus europaeus* and mountain hares L. *timidus* in Sweden. Ecography *30*, 709–715. https://doi.org/10.1111/j.2007.0906-7590.05162.x.
9. Crossman, C.A., Taylor, E.B., and Barrett-Lennard, L.G. (2016). Hybridization in the cetacea: widespread occurrence and associated morphological, behavioral, and ecological factors. Ecol. Evol. *6*, 1293–1303. https://doi.org/10.1002/ece3.1913.
10. Vaz Pinto, P., Beja, P., Ferrand, N., and Godinho, R. (2016). Hybridization following population collapse in a critically endangered antelope. Sci. Rep. *6*, 18788. https://doi.org/10.1038/srep18788.
11. Marshall, T.C., and Spalton, J.A. (2000). Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. Anim. Conserv. *3*, 241–248. https://doi.org/10.1111/j.1469-1795.2000.tb00109.x.
12. Goldberg, T.L., Grant, E.C., Inendino, K.R., Kassler, T.W., Claussen, J.E., and Philipp, D.P. (2005). Increased infectious disease susceptibility resulting from outbreeding depression. Conserv. Biol. *19*, 455–462. https://doi.org/10.1111/j.1523-1739.2005.00091.x.
13. Huff, D.D., Miller, L.M., Chizinski, C.J., and Vondracek, B. (2011). Mixed-source reintroductions lead to outbreeding depression in second-generation descendents of a native North American fish. Mol. Ecol. *20*, 4246–4258. https://doi.org/10.1111/j.1365-294X.2011.05271.x.
14. Rookmaaker, L.C. (1989). The Zoological Exploration of Southern Africa 1650–1790 (A. A. Balkema).
15. Skead, C.J. (1980). Blue antelope (extinct blaauwbok), *Hippotragus leucophaeus* (Pallas, 1766). In Historical Mammal Incidence in the Cape Province (Department of Nature and Environmental Conservation of the Provincial Administration of the Cape of Good Hope), pp. 526–537.
16. Marean, C.W., Cowling, R.M., and 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.
17. Venter, J.A., Brooke, C.F., Marean, C.W., Fritz, H., and Helm, C.W. (2020). Large mammals of the Palaeo-Agulhas Plain showed resilience to extreme climate change but vulnerability to modern human impacts. Quat. Sci. Rev. *235*, 106050. https://doi.org/10.1016/j.quascirev.2019.106050.
18. Cowling, R.M., Potts, A.J., Franklin, J., Midgley, G.F., Engelbrecht, F., and Marean, C.W. (2020). Describing a drowned Pleistocene ecosystem: Last Glacial Maximum vegetation reconstruction of the Palaeo-Agulhas Plain. Quat. Sci. Rev. *235*, 105866. https://doi.org/10.1016/j.quascirev.2019.105866.
19. Jensen, E.L., Díez-del-Molino, D., Gilbert, M.T.P., Bertola, L.D., Borges, F., Cubric-Curik, V., de Navascués, M., Frandsen, P., Heuertz, M., Hvilsom, C., et al. (2022). Ancient and historical DNA in conservation policy. Trends Ecol. Evol. *37*, 420–429. https://doi.org/10.1016/j.tree.2021.12.010.
20. Venier, L.A., and Fahrig, L. (1996). Habitat availability causes the species abundance-distribution relationship. Oikos *76*, 564. https://doi.org/10.2307/3546349.

21. Klein, R.G. (1974). On the taxonomic status, distribution and ecology of the blue antelope, *Hippotragus leucophaeus* (Pallas, 1766). Ann. S. Afr. Mus. *65*, 99–143.

22. Lichtenstein, H. (1811). Reisen im südlichen Africa in den Jahren 1803, 1804, 1805 und 1806 (Berlin C Salfeld).

23. Lichtenstein, H. (1814). Die Gattung Antilope. Ges. naturf. Freunde Berlin Mag. neuesten Entdeck. gesammten Naturk. *6*, 147–182.

24. Klein, R.G. (1983). Palaeoenvironmental implications of Quaternary large mammals in the fynbos region. In Fynbos Paleoecology: A Preliminary Synthesis. South African National Scientific Programmes Rep. Pretoria, H.J. Deacon, Q.B. Hendrey, and J.N.N. Lambrechts, eds. (CSIR), pp. 116–138.

25. Faith, J.T. (2011). Ungulate community richness, grazer extinctions, and human subsistence behavior in southern Africa's Cape Floral Region. Palaeogeogr. Palaeoclimatol. Palaeoecol. *306*, 219–227. https://doi.org/10.1016/j.palaeo.2011.04.025.

26. Cleghorn, N., Potts, A.J., and 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.

27. 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., et al. (2022). Blue turns to gray: paleogenomic insights into the evolutionary history and extinction of the blue antelope (*Hippotragus leucophaeus*). Mol. Biol. Evol. *39*, msac241. https://doi.org/10.1093/molbev/msac241.

28. Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. *15*, e1008302. https://doi.org/10.1371/journal.pgen.1008302.

29. Chardonnet, P., and Crossmary, W. (2013). *Hippotragus equinus* Roan antelope. In Mammals of Africa, J. Kingdon, and M. Hoffmann, eds. (Bloomsbury Publishing), pp. 548–556.

30. Estes, R.D. (2013). *Hippotragus niger* sable antelope. In Mammals of Africa, J. Kingdon, and M. Hoffmann, eds. (Bloomsbury Publishing), pp. 556–565.

31. Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science *348*, 242–245. https://doi.org/10.1126/science.aaa3952.

32. Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L., et al. (2018). Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. Mol. Biol. Evol. *35*, 1225–1237. https://doi.org/10.1093/molbev/msy037.

33. Westbury, M.V., Petersen, B., Garde, E., Heide-Jørgensen, M.P., and Lorenzen, E.D. (2019). Narwhal genome reveals long-term low genetic diversity despite current large abundance size. iScience *15*, 592–599. https://doi.org/10.1016/j.isci.2019.03.023.

34. Dussex, N., van der Valk, T., Morales, H.E., Wheat, C.W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V.E., Guschanski, K., Rhie, A., et al. (2021). Population genomics of the critically endangered kākāpō. Cell Genomics *1*, 100002. https://doi.org/10.1016/j.xgen.2021.100002.

35. Liu, S., Westbury, M.V., Dussex, N., Mitchell, K.J., Sinding, M.-H.S., Heintzman, P.D., Duchêne, D.A., Kapp, J.D., von Seth, J., Heiniger, H., et al. (2021). Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. Cell *184*, 4874–4885.e16. https://doi.org/10.1016/j.cell.2021.07.032.

36. Robinson, J.A., Kyriazis, C.C., Nigenda-Morales, S.F., Beichman, A.C., Rojas-Bracho, L., Robertson, K.M., Fontaine, M.C., Wayne, R.K., Lohmueller, K.E., Taylor, B.L., et al. (2022). The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science *376*, 635–639. https://doi.org/10.1126/science.abm1742.

37. Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D., Lohmueller, K.E., and Wayne, R.K. (2016). Genomic flatlining in the endangered Island fox. Curr. Biol. *26*, 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062.

38. Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E., and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Curr. Biol. *28*, 3487–3494.e4. https://doi.org/10.1016/j.cub.2018.08.066.

39. Waelbroeck, C., Labeyrie, L., Michel, E., Duplessy, J.C., McManus, J.F., Lambeck, K., Balbon, E., and Labracherie, M. (2002). Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. Quat. Sci. Rev. *21*, 295–305. https://doi.org/10.1016/S0277-3791(01)00101-9.

40. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

41. Faith, J.T. (2011). Late Quaternary megafaunal extinctions in southern Africa's Cape Floral Region.

42. Kerley, G.I.H., Sims-Castley, R., Boshoff, A.F., and Cowling, R.M. (2009). Extinction of the blue antelope *Hippotragus leucophaeus*: modeling predicts non-viable global population size as the primary driver. Biodivers. Conserv. *18*, 3235–3242. https://doi.org/10.1007/s10531-009-9639-x.

43. Dupont, L.M., Zhao, X., Charles, C., Faith, J.T., and Braun, D. (2022). Continuous vegetation record of the Greater Cape Floristic Region (South Africa) covering the past 300 000 years (IODP U1479). Clim. Past *18*, 1–21. https://doi.org/10.5194/cp-18-1-2022.

44. Klein, R.G., Avery, G., Cruz-Uribe, K., and Steele, T.E. (2007). The mammalian fauna associated with an archaic hominin skullcap and later Acheulean artifacts at Elandsfontein, Western Cape Province, South Africa. J. Hum. Evol. *52*, 164–186. https://doi.org/10.1016/j.jhevol.2006.08.006.

45. Kaiser, T.M., and Franz-Odendaal, T.A. (2004). A mixed-feeding *Equus* species from the middle Pleistocene of South Africa. Quat. Res. *62*, 316–323. https://doi.org/10.1016/j.yqres.2004.09.002.

46. Stynder, D.D. (2009). The diets of ungulates from the hominid fossil-bearing site of Elandsfontein, Western Cape, South Africa. Quat. Res. *71*, 62–70. https://doi.org/10.1016/j.yqres.2008.06.003.

47. Tyler Faith, J.T., and Thompson, J.C. (2013). Fossil evidence for seasonal calving and migration of extinct blue antelope (*Hippotragus leucophaeus*) in southern Africa. J. Biogeogr. *40*, 2108–2118. https://doi.org/10.1111/jbi.12154.

48. Klein, R.G. (1976). The mammalian fauna of the Klasies River Mouth Sites, Southern Cape Province, South Africa. S. Afr. Archaeol. Bull. *31*, 75. https://doi.org/10.2307/3887730.

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

50. Klingel, H. (2013). *Equus quagga* plains zebra (Common zebra). In Mammals of Africa, J. Kingdon, and M. Hoffmann, eds. (Bloomsbury Publishing), pp. 428–437.

51. van Rensburg, A.P.J. (1975). Die geskiedenis van die Nasionale Bontebokpark, Swellendam. Koedoe *18*, 165–190. https://doi.org/10.4102/koedoe.v18i1.920.

52. Novellie, P., Lindeque, M., Lindeque, P., Lloyd, P., and Koen, J. (2002). Status and action plan for the mountain zebra (*Equus zebra*). IUCN SSC Equid Specialist Group. In Equids: Zebras, Asses and Horses. Status Survey and Conservation Action Plan, P.D. Moehlman, ed. (IUCN). https://portals.iucn.org/library/efiles/documents/2002-043.pdf.

53. Parker, D.M. (2021). Mammals in the mountains: An historical review and updated checklist of the mammals of the Mountain Zebra National Park. Koedoe *63*, a1683. https://doi.org/10.4102/koedoe.v63i1.1683.

54. Hempel, E., Bibi, F., Faith, J.T., Brink, J.S., Kalthoff, D.C., Kamminga, P., Paijmans, J.L.A., Westbury, M.V., Hofreiter, M., and Zachos, F.E. (2021).

Identifying the true number of specimens of the extinct blue antelope (*Hippotragus leucophaeus*). Sci. Rep. *11*, 2100. https://doi.org/10.1038/s41598-020-80142-2.

55. Vrba, E.S., and Gatesy, J. (1994). New antelope fossils from Awash, Ethiopia, and phylogenetic analysis of Hippotragini (Bovidae, Mammalia). Palaeontol. Afr. *31*, 55–72.
56. Deino, A.L. (2011). 40Ar/39Ar dating of Laetoli, Tanzania. In Paleontology and Geology of Laetoli: Human Evolution in Context, T. Harrison, ed. (Springer), pp. 77–97. https://doi.org/10.1007/978-90-481-9956-3_4.
57. Gentry, A.W. (2011). Bovidae. In Paleontology and Geology of Laetoli: Human Evolution in Context, T. Harrison, ed. (Springer), pp. 363–465. https://doi.org/10.1007/978-90-481-9962-4_15.
58. 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. https://doi.org/10.1186/1471-2148-13-166.
59. von Schreber, J.C.D., and Goldfuß, A. (1836). Die Säugthiere in Abbildungen nach der Natur mit Beschreibungen. https://doi.org/10.5962/bhl.title.67399.
60. Barral, D.C., and Seabra, M.C. (2004). The melanosome as a model to study organelle motility in mammals. Pigment Cell Res. *17*, 111–118. https://doi.org/10.1111/j.1600-0749.2004.00138.x.
61. Hume, A.N., Ushakov, D.S., Tarafder, A.K., Ferenczi, M.A., and Seabra, M.C. (2007). Rab27a and MyoVa are the primary Mlph interactors regulating melanosome transport in melanocytes. J. Cell Sci. *120*, 3111–3122. https://doi.org/10.1242/jcs.010207.
62. Wu, X.S., Masedunskas, A., Weigert, R., Copeland, N.G., Jenkins, N.A., and Hammer, J.A. (2012). Melanoregulin regulates a shedding mechanism that drives melanosome transfer from melanocytes to keratinocytes. Proc. Natl. Acad. Sci. USA *109*, E2101–E2109. https://doi.org/10.1073/pnas.1209397109.
63. Ambrosio, A.L., Boyle, J.A., Aradi, A.E., Christian, K.A., and Di Pietro, S.M. (2016). TPC2 controls pigmentation by regulating melanosome pH and size. Proc. Natl. Acad. Sci. USA *113*, 5622–5627. https://doi.org/10.1073/pnas.1600108113.
64. Brunberg, E., Andersson, L., Cothran, G., Sandberg, K., Mikko, S., and Lindgren, G. (2006). A missense mutation in PMEL17 is associated with the silver coat color in the horse. BMC Genet. *7*, 46. https://doi.org/10.1186/1471-2156-7-46.
65. Runkel, F., Büssow, H., Seburn, K.L., Cox, G.A., Ward, D.M., Kaplan, J., and Franz, T. (2006). Grey, a novel mutation in the murine Lyst gene, causes the beige phenotype by skipping of exon 25. Mamm. Genome *17*, 203–210. https://doi.org/10.1007/s00335-005-0015-1.
66. Li, W., Sartelet, A., Tamma, N., Coppieters, W., Georges, M., and Charlier, C. (2016). Reverse genetic screen for loss-of-function mutations uncovers a frameshifting deletion in the melanophilin gene accountable for a distinctive coat color in Belgian Blue cattle. Anim. Genet. *47*, 110–113. https://doi.org/10.1111/age.12383.
67. Thapa, P.C., Do, D.N., Manafiazar, G., and Miar, Y. (2023). Coat color inheritance in American mink. BMC Genom. *24*, 234. https://doi.org/10.1186/s12864-023-09348-8.
68. Hoekstra, H.E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity (Edinb) *97*, 222–234. https://doi.org/10.1038/sj.hdy.6800861.
69. Ng, P.C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814. https://doi.org/10.1093/nar/gkg509.
70. Yang, C.W., Ran, J.S., Yu, C.L., Qiu, M.H., Zhang, Z.R., Du, H.R., Li, Q.Y., Xiong, X., Song, X.Y., Xia, B., et al. (2019). Polymorphism in MC1R, TYR and ASIP genes in different colored feather chickens. 3 Biotech *9*, 203. https://doi.org/10.1007/s13205-019-1710-z.
71. Barsh, G.S. (1996). The genetics of pigmentation: from fancy genes to complex traits. Trends Genet. *12*, 299–305. https://doi.org/10.1016/0168-9525(96)10031-7.
72. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.
73. Gansauge, M.T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., and Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. *45*, e79. https://doi.org/10.1093/nar/gkx033.
74. Gansauge, M.T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. *8*, 737–748. https://doi.org/10.1038/nprot.2013.038.
75. Paijmans, J.L.A., Baleka, S., Henneberger, K., Taron, U.H., Trinks, A., Westbury, M.V., and Barlow, A. (2017). Sequencing single-stranded libraries on the Illumina NextSeq 500 platform. Preprint at arXiv. https://doi.org/10.48550/arXiv.1711.11004.
76. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
77. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta *2*, e107. https://doi.org/10.1002/imt2.107.
78. Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*, R46. https://doi.org/10.1186/gb-2014-15-3-r46.
79. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082. https://doi.org/10.1038/nprot.2014.063.
80. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
81. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
82. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media, Inc.).
83. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
84. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
85. Lin, M.F., Rodeh, O., Penn, J., Bai, X., Reid, J.G., Krasheninina, O., and Salerno, W.J. (2018). GLnexus: joint variant calling for large cohort sequencing. Preprint at bioRxiv. https://doi.org/10.1101/343970.
86. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.
87. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at arXiv. https://doi.org/10.48550/arXiv.1207.3907.
88. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
89. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinform. *15*, 356. https://doi.org/10.1186/s12859-014-0356-4.

90. Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M.V., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366*, 594–599. https://doi.org/10.1126/science.aaw2090.

91. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

92. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w ^{1118}; iso-2; iso-3. Fly *6*, 80–92. https://doi.org/10.4161/fly.19695.

93. Renaud, G., Hanghøj, K., Willerslev, E., and Orlando, L. (2017). Gargammel: a sequence simulator for ancient DNA. Bioinformatics *33*, 577–579. https://doi.org/10.1093/bioinformatics/btw670.

94. Rabiee, M., Sayyari, E., and Mirarab, S. (2019). Multi-allele species reconstruction using ASTRAL. Mol. Phylogenet. Evol. *130*, 286–296. https://doi.org/10.1016/j.ympev.2018.10.033.

95. García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics *28*, 2678–2679. https://doi.org/10.1093/bioinformatics/bts503.

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

97. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

98. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226. https://doi.org/10.1126/science.1224344.

99. Koepfli, K.P., Tamazian, G., Wildt, D., Dobrynin, P., Kim, C., Frandsen, P.B., Godinho, R., Yurchenko, A.A., Komissarov, A., Krasheninnikova, K., et al. (2019). Whole genome sequencing and re-sequencing of the sable antelope (*Hippotragus niger*): a resource for monitoring diversity in *ex situ* and *in situ* populations. G3 (Bethesda) *9*, 1785–1793. https://doi.org/10.1534/g3.119.400084.

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

101. Cahill, J.A., Heintzman, P.D., Harris, K., Teasdale, M.D., Kapp, J., Soares, A.E.R., Stirling, I., Bradley, D., Edwards, C.J., Graim, K., et al. (2018). Genomic evidence of widespread admixture from polar bears into brown bears during the Last Ice Age. Mol. Biol. Evol. *35*, 1120–1129. https://doi.org/10.1093/molbev/msy018.

102. Oliva, A., Tobler, R., Cooper, A., Llamas, B., and Souilmi, Y. (2021). Systematic benchmark of ancient DNA read mapping. Brief. Bioinform. *22*, bbab076. https://doi.org/10.1093/bib/bbab076.

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

104. Xu, W., Lin, Y., Zhao, K., Li, H., Tian, Y., Ngatia, J.N., Ma, Y., Sahu, S.K., Guo, H., Guo, X., et al. (2021). An efficient pipeline for ancient DNA mapping and recovery of endogenous ancient DNA from whole-genome sequencing data. Ecol. Evol. *11*, 390–401. https://doi.org/10.1002/ece3.7056.

105. Humble, E., Dobrynin, P., Senn, H., Chuven, J., Scott, A.F., Mohr, D.W., Dudchenko, O., Omer, A.D., Colaric, Z., Lieberman Aiden, E.L., et al. (2020). Chromosomal-level genome assembly of the scimitar-horned oryx: insights into diversity and demography of a species extinct in the wild. Mol. Ecol. Resour. *20*, 1668–1681. https://doi.org/10.1111/1755-0998.13181.

106. Gonçalves, M., Siegismund, H.R., Jansen van Vuuren, B., Koepfli, K.P., Ferrand, N., and Godinho, R. (2021). *De novo* whole-genome assembly and resequencing resources for the roan (*Hippotragus equinus*), an iconic African antelope. G3: Genes, Genomes, Genet. *11*, jkab002. https://doi.org/10.1093/g3journal/jkab002.

107. Prasad, A., Lorenzen, E.D., and Westbury, M.V. (2022). Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. Mol. Ecol. Resour. *22*, 45–55. https://doi.org/10.1111/1755-0998.13457.

108. IUCN SSC Antelope Specialist Group (2017). *Hippotragus niger*. The IUCN Red List of Threatened Species, e.T10170A50188654. https://doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10170A50188654.en.

109. IUCN SSC Antelope Specialist Group (2017). *Hippotragus equinus*. The IUCN Red List of Threatened Species, e.T10167A50188287. https://doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10170A50188654.en.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
QIAGEN MinElute kit	QIAGEN	Cat#28004
PippinPrep Casette 3% agarose, w/EtBr, external marker C, 90-250 bp	SageScience	Cat#CSD3010
Critical commercial assays		
D1000 Screen Tape (Tapestation 2200)	Agilent	Cat#5067-5583
dsDNA HS Assay Kit (Qubit 2.0)	ThermoFisher	Cat#Q32851
Deposited data		
Blue antelope raw sequence data, fastq format, SRR24014091–SRR24014111	this study	N/A
Blue antelope raw sequence data, fastq format, SRR20324702	Hempel et al.[54]	N/A
Blue antelope raw sequence data, fastq format, SRR18753915, SRR18936629–SRR18936631	Hempel et al.[27]	N/A
Oligonucleotides		
IS7 amplification primer: ACACTCTTTCCCTACACGAC	Meyer and Kircher[72]	Sigma Aldrich
IS8 amplification primer: GTGACTGGAGTTCAGACGTGT	Meyer and Kircher[72]	Sigma Aldrich
IS5 re-amplification primer: AATGATACGGCGACCACCGA	Meyer and Kircher[72]	Sigma Aldrich
IS6 amplification primer: CAAGCAGAAGACGGCATACGA	Meyer and Kircher[72]	Sigma Aldrich
Splinter Oligo: TL93: AmC12-AACTTCCGATCTNNNNNN-AmC7, TL106: AmC12-AACTTCCGATCTNNNNNN-AmC3, TL110: SpacerC12-AA[SpacerC12]CTTCCG ATCTNNNNNN-AmC6, TL136: SpacerC12-AA[SpacerC12]CTTCCG ATCTNNNNNNN-AmC6, TL137: SpacerC12-AA[SpacerC12]CTTCCGA TCTNNNNNNNN-AmC6 (100 μM)	Gansauge et al.[73]	Sigma Aldrich
CL9 extension primer: GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT (100 μM)	Gansauge and Meyer[74]	Sigma Aldrich
Double-stranded adaptor Strand 1 (CL53): CGACGCTCTTC-ddC (ddC = dideoxycytidine) (500 μM) Strand 2 (CL73): [Phosphate]GGAA GAGCGTCGTGTAGGGAAA GAG*T*G*T*A (* = phosphothioate linkage) (10 μM)	Gansauge and Meyer[74]	Sigma Aldrich
CL78 single-stranded adapter: AGAT CGGAAG[C3Spacer] $_{10}$ [TEG-biotin] (TEG = triethylene glycol spacer) (10 μM)	Gansauge and Meyer[74]	Sigma Aldrich
CL130 GTGACTGGAGTTCAG ACGTGTGCTCTTCC*GA*TC*T	Gansauge et al.[73]	Sigma Aldrich
CL72 R1 sequencing primer: ACACTCTTTCCCTACACGACGCTCTTCC	Gansauge and Meyer[74]	Sigma Aldrich
Gesaffelstein index 2 sequencing primer: GGAAG AGCGTCGTGTAGGGAAAGAGTGT	Paijmans et al.[75]	Sigma Aldrich
P5 indexing primer: AATGATACGGCGACC ACCGAGATCTACACnnnnnnnnACACTCTT TCCCTACACGACGCTCTT	Gansauge and Meyer[74]	Sigma Aldrich

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
P7 indexing primer: CAAGCAGAAGACGGC ATACGAGATnnnnnnnnGTGACTGGAGTTCA GACGTGT	Gansauge and Meyer[74]	Sigma Aldrich
Software and algorithms		
fastp v0.23.2/v0.23.4	Chen et al.[76]; Chen[77]	https://github.com/OpenGene/fastp
Kraken v2.1.3	Wood and Salzberg[78]	https://ccb.jhu.edu/software/kraken
PALEOMIX v1.3.8	Schubert et al.[79]	https://paleomix.readthedocs.io/en/stable
bwa aln v0.7.17	Li and Durbin[80]	https://github.com/lh3/bwa
samtools v1.16.1/v1.18	Li et al.[81]	http://www.htslib.org
GATK4 v4.4.0.0	Van der Auwera and O'Connor[82]	https://gatk.broadinstitute.org/hc/en-us
Picard Toolkit v2.27.4		http://broadinstitute.github.io/picard
SAMtools v1.16.1	Li et al.[81]	
bcftools v1.16	Danecek et al.[83]; Li[84]	https://samtools.github.io/bcftools/
GLnexus v1.4.1	Lin et al.[85]	https://github.com/dnanexus-rnd/GLnexus
mapDamage v2.2.1	Jónsson et al.[86]	https://ginolhac.github.io/mapDamage/
freebayes v0.9.21.7	Garrison and Marth[87]	https://github.com/freebayes/freebayes
IQ-Tree2 v2.2.0.3	Minh et al.[88]	http://www.iqtree.org
PSMC v0.6.5	Li and Durbin[40]	https://github.com/lh3/psmc
ANGSD v0.923	Korneliussen et al.[89]	https://github.com/ANGSD/angsd
QuiBL	Edelman et al.[90]	https://github.com/miriammiyagi/QuIBL
bedtools v2.29.1	Quinlan and Hall[91]	https://bedtools.readthedocs.io/en/latest/
SnpEff/SnpSift v5.1/v5.2	Cingolani et al.[92]	http://pcingola.github.io/SnpEff/,
gargammel v1.1.2	Renaud et al.[93]	https://grenaud.github.io/gargammel/
ASTRAL v5.7.8	Rabiee et al.[94]	https://github.com/smirarab/ASTRAL
QualiMap v2.3	García-Alcalde et al.[95]	http://qualimap.conesalab.org/
DamageProfiler v1.1	Neukamm et al.[96]	https://github.com/Integrative-Transcriptomics/DamageProfiler
MultiQC v1.18	Ewels et al.[97]	https://multiqc.info/
Basecalls_0.1.sh	this study	https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780)
ROH.py	this study	https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780)
mutator.R	this study	https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780)
countSNPs.pl	this study	https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780)
findTracts.pl	this study	https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780)
Other		
SYBR Green PCR MasterMix	Thermo Fisher	Cat#4309155
AccuPrime™ Pfx DNA Polymerase	Thermo Fisher	Cat#12344024
T4 DNA Ligase (5 U/mL)	Thermo Fisher	Cat#EL0011
T4 DNA Ligase (30 U/mL)	Thermo Fisher	Cat#EL0013
FastAP Thermosensitive Alkaline Phosphatase (1 U/μL)	Thermo Fisher	Cat#EF0651
PEG-4000 (50% (w/v))	Thermo Fisher	Cat#EL0011

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HPLC grade water	A. Hartenstein	Cat#CW20
Tween 20 100%	A. Hartenstein	Cat#CT21
EDTA Solution, 0.5M	VWR	Cat#E177-500MLDB
T4 RNA Ligase Buffer & PEG 8000	New England BioLabs	Cat#B0216L
ATP 100mM	Thermo Fisher	Cat#R0441
Tris-HCl solution, 1M pH8	Thermo Fisher	Cat#15568-025
NaCl solution, 5M	Thermo Fisher	Cat#AM9760G
Dynabeads My One Streptavidin C1	Thermo Fisher	Cat#65001
Klenow fragment of DNA polymerase I	Thermo Fisher	Cat#EP0051
dNTP 25mM each Mix	Biozym	Cat#350600502 (BR0600502)
SDS solution, 20%	Thermo Fisher	Cat#AM9820
SSC Buffer, 20x	Roth	Cat#1054.1
TE Buffer	VWR	Cat#E112-100ml
Nuclease-free Water	New England BioLabs	Cat#B1500S
USER Enzyme	New England BioLabs	Cat#M5505S
Herculase II Fusion DNA Polymerase	Agilent	Cat#600677
T4 PNK	Thermo Fisher	Cat#EK0031

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Prof. Dr. Michael Hofreiter (michael.hofreiter@uni-potsdam.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The untrimmed single-stranded library raw data has been deposited at the Sequence Read Archive and is publicly available as of the date of publication. Accession numbers are listed in the key resources table. The data is associated in GenBank with BioProject number PRJNA950236. The BioSample is SAMN22745443, the same as in Hempel et al.[54] In addition, publicly available data from Hempel et al.[54] and Hempel et al.[27] was used. The accession numbers for the datasets are listed in the key resources table. All original code is available on https://github.com/Mvwestbury/Blue_antelope_tools (https://doi.org/10.5281/zenodo.10780). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We used the extract and a single-stranded library of the historical blue antelope (*Hippotragus leucophaeus*) mounted skin (NRM 590107) housed in the Swedish Museum of Natural History (Naturhistoriska riksmuseet) in Stockholm from a previous study.[54] From the extract we built an additional eight libraries (see below). This specimen does not have locality information.

METHOD DETAILS

Lab procedures

DNA preparation

We built eight single-stranded libraries using 1.73 μl of an already existing DNA extract for each.[54,73] We performed uracil removal through an initial 15 min incubation step with 0.5 μl USER (Uracil-Specific Excision Reagent) Enzyme (modified from Meyer et al.[98]). The maximum amount of input DNA used for library construction was 10 ng. A qPCR (Thermo Scientific PikoReal Real-Time PCR System) was run to determine the optimal number of amplification cycles for the subsequent dual-indexing PCR. Extraction and library blanks were run alongside to check for contamination. Since the short read length of this sample was already known from a previous study,[54] we removed very short reads to improve the mappable proportion of the sequenced reads by processing all eight newly built libraries on the Pippin Prep device (Sage Science) with a 3% Agarose Gel Cassette for targets from 90–250 bp (Marker C) with selection parameters set to 165–300 bp. Subsequently all libraries were re-amplified using Herculase II Fusion Polymerase. The

library used in Hempel et al.[54] had been treated in the same way. All pre-PCR lab work was conducted in dedicated archival DNA facilities at the University of Potsdam, Germany. The eight newly constructed libraries were sequenced using custom primers in two runs on a NovaSeq6000 at the SciLifeLab, Sweden, generating 101 bp paired-end reads.[74,75] The original library was sequenced on one of these runs. In addition, four of the libraries were sequenced using custom primers on an Illumina NextSeq500 at University of Potsdam, Germany, generating 75 bp single-end reads. We also used 75 bp single-end read data from NRM 590107 from two previous studies generated from the original library[27,54] (Table S1).

Bioinformatic procedures

aDNA simulation

To determine the best mapping approach in terms of data mapped and accuracy of downstream variant calling, we simulated 150 bp paired-end Illumina sequencing reads from the largest scaffold of *H. niger* (Genbank code: VHQK01014937.1[99]; 19,097,140 bp) using gargammel v1.1.2[93] to a coverage of ~30x. To evaluate the impact of read length and/or damage, we simulated three datasets: i) 300 bp fragment lengths with no damage, ii) Read lengths based on the output from mapDamage for the blue antelope without damage, and iii) Read lengths and C-T and A-G damage patterns based on the output from mapDamage. For iii) we subsequently randomly mutated G into T across the read at an ~0.33% rate using a custom R script (mutator.R, https://github.com/Mvwestbury/Blue_antelope_tools). For each of the three datasets, we ran seven different mapping approaches based on previously published parameters for mapping ancient DNA data[100–104] (Table S2).

Data processing and read mapping

We treated single- and paired-end reads separately but combined the reads of each library prior to duplicate removal. We used fastp v0.23.2,[76,77] https://github.com/OpenGene/fastp) to trim Illumina adapter sequences, remove reads <30 bp (--length-required) and merge paired-end reads (--merge) with disabled quality filtering (-Q) and trimming poly G tails (--trim_poly_g). We mapped the resulting reads to scaffolds 1-29 (https://www.ncbi.nlm.nih.gov/assembly/GCF_014754425.289[105]) and the mitochondrial genome of the same individual of the scimitar-horned oryx *(Oryx dammah)* using the bwa aln algorithm v0.7.17[80] with a maximum edit distance (-n) of 0.01 and a maximum number of gap openings (-o) of 2 while using only merged reads for paired-end data. We removed reads with a mapping quality of <30 and discarded unmapped reads with SAMtools view v1.16.1.[81] We merged all bam files originating from the same library with SAMtools merge and sorted them with SAMtools sort. Subsequently, we removed duplicates with Picard MarkDuplicates v2.27.4 (Picard Toolkit 2020 — http://broadinstitute.github.io/picard). We sorted and merged all bam files into one single file for sample NRM 590107 and sorted it again. Lastly, we performed a mapDamage v2.2.0[86] analysis (Figure S1) and rescaled the bam file to decrease the quality of misincorporations that are likely caused by aDNA damage using the rescale option (--rescale).

We also used published raw sequencing read data from contemporary samples of captive specimens of the roan (*H. equinus*; sample 10954[106]) and the sable antelope (*H. niger*; sample SB2152[99]). The reads were processed in the same way as described for the blue antelope and mapped to the same reference genome using the bwa mem algorithm (instead of bwa aln) with default settings and both merged and unmerged reads. For the roan antelope sample, which originates from a 10X Genomics library, 22 bp were trimmed from R1 while using fastp (--trim_front1=22). Sex chromosomes as well as mitochondrial genomes are inherited differently compared to autosomes. To avoid biasing downstream analyses, we removed scaffold 23 which was previously shown to align best with the X chromosome in the cattle genome.[105]

We chose the scimitar-horned oryx as the reference genome due to its relative high quality and relatively close relationship to the genus *Hippotragus*. We mapped all *Hippotragus* species to the outgroup so all individuals will be equidistant to the reference genome helping to alleviate some biases that can be caused by unequal phylogenetic distances to the reference genome in genetic diversity and demographic analyses.[107] More specifically, it was shown that with increased phylogenetic distance to the reference, depending on filtering parameter choice, there are increasing levels of heterozygosity and as a result decreasing levels of runs of homozygosity.[107] However, as the scimitar-horned oryx is only ~4 million years divergent, we do not expect the biases caused by phylogenetic distance of the reference genome to be large as it was shown previously that even at a divergence of ~7 million years, the impact is minor.[107]

QUANTIFICATION AND STATISTICAL ANALYSIS

Estimation of effective population size

We used freebayes v0.9.21.7[87] to generate filtered vcf files as input for the pairwise sequential Markovian coalescent (PSMC) model analyses. We first generated vcf files from the rescaled blue antelope bam file and the roan and the sable antelope bam files in which we called heterozygous positions when at least 30% showed the minor allele using the --min-alternate-fraction option while setting the minimum coverage to half the average coverage (--min-coverage). This resulted in a minimum coverage value of 21 for the blue and the roan antelope and 16 for the sable antelope. We also filtered for base and mapping quality (--min-mapping-quality 20, --min-base-quality 20). We then repeated this calling of heterozygous positions using default parameters (at least 5% showed the minor allele). Subsequently, we used bcftools isec v1.16 to create a complement vcf file with sites that are only present in the 5% file (-p -C).[83] We then removed all indels from this file and created a list with SNPs that are only present in the 5% but not in the 30% file to exclude from the analysis. To then create a vcf file without heterozygous SNPs with a minor allele proportion of <0.3 but also including non-variant base calls which is necessary for the PSMC analysis, we used bcftools mpileup (-Ou -f --annotate FORMAT/DP), bcftools call (-c)

and bcftools view with the -T option.[84] Only autosomes were used in the analysis. We generated a diploid consensus sequence from the vcf files using vcfutils (minimum read depth -d 14, maximum read depth -D 87).[83] Next, we generated the PSMC fasta input file with fq2psmcfa using a quality cutoff of 20 (-q). We then inferred changes in effective population size (N_e) through time with PSMC v0.6.5[40] with parameters previously shown to be relevant with human data (-N25 -t15 -r5 -p "4 + 25 × 2 + 4+6"). To perform bootstrapping analyses using 100 replicates we used splitfa (from the psmc package) to cut long sequences of the input file and to sample randomly with replacement. For plotting the blue antelope data we used the generation time of the sable (7.1 years[108]), the closest relative of the blue antelope,[27] and a per-generation mutation rate of 1.14 x 10^{-8} based on the yearly per site mutation rate (see below). In addition, we performed PSMC analyses for the roan and the sable antelope with adjusted parameters for vcfutils (roan: minimum read depth -d 14, maximum read depth -D 88; sable: minimum read depth -d 11, maximum read depth -D 67). For plotting we used a generation time of 7.5 years for the roan antelope[109] and per-generation mutation rates of 1.19 x 10^{-8} for the sable and 1.26 x 10^{-8} for the roan antelope.

We inferred the mutation rate per site per year for each of the three *Hippotragus* species using the data mapped to the scimitar-horned oryx. We determined the genetic distance between each species and the outgroup species using only autosomes in ANGSD v0.923 (-doIBS 2 -doCounts 1 -makeMatrix 1 -minMapQ 20 -minQ 20 -minInd 2).[89] Using the split from our fossil-calibrated phylogeny of 4.03 Mya between blue/sable/roan antelope and scimitar-horned oryx, we inferred the mutation rate per site per year to be 1.61 x 10^{-9} for the blue, 1.68 x 10^{-9} for the sable and 1.69 x 10^{-9} for the roan antelope.

Heterozygosity

We used ANGSD to directly perform allele counts on the autosomes of the mapped bam files, including the rescaled blue antelope, using the following command and filters: -minQ 20 -minMapQ 20 -uniqueOnly 1 -remove_bads 1 -doCounts 1 -dumpCounts 4 -setMinDepthInd 10. We wrote a custom script (Basecalls_0.1.sh, https://github.com/Mvwestbury/Blue_antelope_tools) to only call heterozygous sites if the minor allele was >10% of all alleles. We plotted all resultant minor allele frequencies for heterozygous base calls (Figure S2). Visual inspection showed that a 30% threshold would accurately return most heterozygous sites in the modern individuals. Therefore, we reran the analysis with a 30% threshold (Basecalls_0.3.sh). A site with an allele found at a frequency of >70% was called homozygous. Any sites that did not have two alleles with a frequency of >30% each or one allele with >70% frequency was counted but not given a base call. We divided the total number of heterozygous bases by the total number of sites that passed the initial ANGSD filtering in the individual.

Runs of homozygosity

Using the results from the heterozygosity analysis, we counted runs of homozygosity in 100 kb windows using a custom script (ROH.py, https://github.com/Mvwestbury/Blue_antelope_tools). We only considered a window as ROH if the proportion of heterozygous sites in that window was <0.0001 and the window contained at least 25 kb of called sites. We further counted the number of consecutive 100 kb windows in ROH (Figure S3) and obtained the number of consecutive windows >1 Mb in length.

Genetic load

We calculated the genetic load in the three species using SnpEff v5.1 (http://pcingola.github.io/SnpEff[92]) using the vcf file with a 30% minor allele cutoff from freebayes. We built the SnpEff database using the config file available with the software but with the addition of the scimitar-horned oryx genome[105] and annotations using the build function. We ran SnpEff independently for each vcf file and filtered the annotated output vcf based on two groupings. A SNP was designated loss of function if the site was annotated as either "stop_gained", "frameshift_variant", "splice_acceptor_variant", or "splice_donor_variant". A SNP was designated as synonymous if a site was annotated as either "synonymous_variant", "start_retained", or "stop_retained_variant". When investigating realized genetic load, we only considered homozygous sites when the individual of interest had the derived allele — the reference allele was considered the ancestral allele as we mapped to the outgroup.

Fossil calibrated phylogenies

For the sliding window tree phylogeny we generated pseudohaploid fasta files using the most common base as input (-doFasta 2) in ANGSD[89] with the following parameters: -doCounts 1, -minQ 20, -minMapQ 20, -uniqueonly 1, -remove_bads 1, -only_proper_pairs 0, -trim 0, -setMinDepthInd 5, -doMajorMinor 1, -GL 2. We only included autosomes in the analyses. As input files we used the rescaled bam files from mapDamage. From the fasta files we generated 20 kb sliding windows with a 1 Mb slide using bedtools v2.29.1[91] and the makewindows function to generate the coordinates for the windows and SAMtools faidx to extract the sequence. In R we removed windows when any of the individuals had more than 50% missing data. We ran each window through IQ-Tree v2.2.0.3[88] using the parameters -st DNA -m GTR+R6 -bb 1000 -alrt 0. We calculated the frequency of each possible quartet using ASTRAL v5.7.8[94] and the -t 32 parameter. For dating, we further filtered the sliding windows into three different sets, one for each of the three potential topologies. The fossil-calibrated phylogenies were generated as described in Hempel et al.[27]

Gene flow

D *statistics*

We performed a *D* statistics test for gene flow with a random base call approach using only autosomes (-doabbababa 1) in ANGSD and the following parameters -minQ 30 -minMapQ 30 -setMinDepthInd 5 -uniqueOnly 1. We specified the scimitar-horned oryx as the

outgroup using -uselast 1. To test for significance we performed a block jackknifing approach in 5 Mb windows using the ANGSD_jackknife.R script which is part of the ANGSD toolsuite.

SNP tracts

The consensus sequences obtained as described for the sliding window tree analysis were used to identify, with two custom Perl scripts (countSNPs.pl and findTracts.pl, https://github.com/Mvwestbury/Blue_antelope_tools), all SNP positions that correspond to the species tree [((blue, sable), roan), outgroup)] or the alternative topologies [(((sable, roan), blue), outgroup); (((blue, roan), sable), outgroup)]. Testing and manual inspection of different patterns of consecutive, uninterrupted SNPs were used and evaluated. By means of these results, we used the following criteria to differentiate between signal and noise: at least ten consecutively scored sites of a focal topology were extracted, but up to four consecutive sites of the other two topologies were allowed. From these, only tracts in which 75% of the positions were scored as the focal topology were kept. The start and end positions of these tracts were then extracted from the genome.

QuIBL

To test hypotheses of whether phylogenetic discordances can be explained by incomplete lineage sorting (ILS) alone, or by a combination of ILS and gene flow, we applied QuIBL[90]. As input files we used the gene trees from the sliding window phylogenomic analysis. We ran QuIBL specifying the scimitar-horned oryx as the overall outgroup (totaloutgroup), to test either ILS or ILS with gene flow (numdistributions 2), the number of total EM steps as 50 (numsteps), and a likelihood threshold of 0.01. We determined the significance of gene flow by comparing the BIC1 (ILS alone) and BIC2 (ILS and gene flow). When BIC2 was lower than BIC1, with a difference of >10, we assumed incongruent topologies arose due to both ILS and gene flow.

Coat color analysis

Variant calling against roan antelope assembly

Sequencing reads were trimmed using fastp v0.23.4 to remove low quality (< 25, -q) ends of reads, remove read lengths under 35 bp (-l), merge overlapping paired-end reads (-m) by minimum overlap of 15 bp (--overlap_len_require) and maximum difference of 3 bp (--overlap_diff_limit), enabled base correction on overlapping regions (-c), polyG tail trimming (-g) and polyX trimming in 3' ends (-x). Sequence contaminants were detected using Kraken v2.1.3[78] with a confidence of 0.8 and the pre-compiled 8 GB database constructed from complete bacterial, archaeal, and viral genomes in RefSeq (2017-10-18). Unclassified trimmed reads were aligned to the roan antelope (*Hippotragus equinus*) reference genome assembly GCA_016433095.1[106] using the bwa aln algorithm for merged reads only with a seed of 16,500, maximum distance of 0.01 (-n), and a maximum number of gap opens of 2 (-o). Duplicate reads were marked using PALEOMIX v1.3.8[79] for bwa aln aligned reads. Bam files from the same sample generated by multiple runs were merged using SAMtools v1.18. Base recalibration was done using mapDamage. Alignment quality was assessed using QualiMap v2.3,[95] DamageProfiler v1.1,[96] and MultiQC v1.18.[97] Germline variants were detected using both mpileup calling in SAMtools and BCFtools v1.18[83,84] and HaplotypeCaller in GATK4 v4.4.0.0.[82] Genotyping of gvcf files was determined using GLnexus v1.4.1.[85] Variant effects were annotated using SnpEff and SnpSift v5.2 with a custom SIFT database built for the roan antelope.

Candidate genes for color and patterning

We defined the coloration phenotype of interest for the blue antelope — its purportedly blue or gray coat color — as eumelanin dilution, which refers to any lighter shade appearance of eumelanin pigment-based coloration. We examined a non-exhaustive list of 49 candidate genes connected to eumelanin dilution (*MLPH*, *RAB27A*, *MYO5A*, *ABCB6*, *AP3B1*, *AP3D1*, *ATP7A*, *BCL2*, *BLOC1S1*, *BLOC1S2*, *BLOC1S3*, *BLOC1S4*, *BLOC1S5*, *BLOC1S6*, *DCT*, *DTNBP1*, *EED,* FIG4, *GPR143*, *HPS1*, *HPS2*, *HPS3*, *HPS4*, *HPS5*, *HPS6*, *LYST*, *MBTPS1*, *MCOLN3*, *MFSD12*, *MKLN1*, *MLANA*, *MREG*, *TPCN2*, *PAH*, *PMEL*, *RAB38*, *RABGGTA*, *SLC24A5*, *SLC45A2*, *SLC7A11*, *SNAI2*, *SPAG9*, *VAC14*, *VPS33*) or having well-documented roles in mammalian color and patterning (*MITF*, *ASIP*, *MC1R*, *TYR*, *TYRP1*) for sequence variants detected in the blue antelope as homozygous and non-synonymous relative to the roan antelope, and which also had low SIFT scores (Table S4).