A highly divergent mitochondrial genome in extant Cape buffalo from Addo Elephant National Park, South Africa

Deon de Jager[1]*, Marlo Möller[2,3], Eileen Hoal[2], Paul van Helden[2], Brigitte Glanzmann[2], Cindy Harper[4], Paulette Bloomer[1]

[1]Molecular Ecology and Evolution Programme, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

[2]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town.

[3]Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa

[4]Veterinary Genetics Laboratory, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa

*Corresponding author: Deon de Jager, dejager4@gmail.com

Abstract

The reduced costs of next-generation sequencing (NGS) has allowed researchers to generate nuclear and mitochondrial genome data to gain deeper insights into the phylogeography, evolutionary history, and biology of non-model species. While the Cape buffalo (*Syncerus caffer caffer*) has been well-studied across its range with traditional genetic markers over the last 25 years, researchers are building on this knowledge by generating whole genome, population-level datasets to improve understanding of its genetic composition and evolutionary history. Using publicly available NGS data, we assembled 40 Cape buffalo mitochondrial genomes (mitogenomes) from four protected areas in South Africa, expanding the geographical range and almost doubling the number of mitogenomes available for this species. Coverage of the mitogenomes ranged from 154-1,036X. Haplotype and nucleotide diversity for Kruger National Park (n = 15) and Mokala National Park (n = 5) were similar to diversity levels in southern and eastern Africa. Hluhluwe-iMfolozi Park (n = 15) had low levels of genetic diversity, with only four haplotypes detected, reflecting its past bottleneck. Addo Elephant National Park (n = 5) had the highest nucleotide diversity of all populations across Africa, which was unexpected, as it is known from previous studies to have low nuclear diversity. This diversity was driven by a highly divergent mitogenome from one sample, which was also identified in another sample via Sanger sequencing of cytochrome *b*. Using a fossil-calibrated phylogenetic analysis, we estimated that this mitogenome diverged from all other Cape buffalo approximately 2.15 million years ago. We discuss several potential sources of this mitogenome but propose that it most likely originated through introgressive hybridisation with the extinct long-horned buffalo, *Syncerus* (*Pelorovis*) *antiquus*. We conclude by discussing the conservation consequences of this finding for the Addo Elephant National Park population, proposing careful genetic management to prevent inbreeding depression while maintaining this highly unique diversity.

Introduction

The Cape buffalo (*Syncerus caffer caffer*, Sparrman 1779) is a charismatic African bovid known for being one of the "Big Five". It is also a reservoir for important veterinary diseases, such as foot-and-mouth disease, Corridor disease, brucellosis, and bovine tuberculosis (Laubscher & Hoffman, 2012), and is a high-value species in both hunting and wildlife ranching, particularly in South Africa (Taylor, Lindsey, & Davies-Mostert, 2016). Consequently, this species has been well-studied from ecological, disease, and genetic perspectives throughout most of its range (Cornélis et al., 2014; de Jager et al., 2021; de Jager, Harper, & Bloomer, 2020; Heller, Lorenzen, Okello, Masembe, & Siegismund, 2008; Heller, Okello, & Siegismund, 2010; O'Ryan et al., 1998; Simonsen, Siegismund, & Arctander, 1998; Smitz et al., 2013; Smitz et al., 2014; van Hooft, Groen, & Prins, 2000, 2002, 2003).

However, most of the genetic studies have been based on microsatellites, which are difficult to combine and compare between studies, or on a portion of the control region of the mitochondrial genome (mitogenome), which is easier to combine across studies, but can lack the power to detect deep lineages within species. With the ever-decreasing cost of next-generation sequencing, it is possible to generate population-level whole genome sequences for non-model species, as illustrated for Cape buffalo in South Africa by de Jager et al. (2021). In their study, Heller, Brüniche-Olsen, and Siegismund (2012) generated population-level mitogenomes from populations in Ethiopia, Kenya, Zimbabwe, and Botswana. These types of data are easier to combine across studies and are more powerful than traditional nuclear and mitochondrial markers. However, some gaps in genomic data still exist, for example the lack of nuclear genomes from the rest of southern Africa and eastern Africa, and the lack of mitogenomes from South Africa.

In the latest Red List assessment by the International Union for Conservation of Nature (IUCN), the African buffalo (*S. caffer*) was downgraded from Least Concern to Near Threatened, predominantly due to ongoing and predicted population declines (IUCN, 2019). However, Cape buffalo is regionally classified as Least Concern in southern Africa and is increasing in numbers in South Africa (Tambling, Venter, du Toit, & Child, 2016). In this study, we present 40 Cape buffalo mitogenomes from four South African protected areas to fill this gap in mitogenomic data for the species. We analyse these in the context of previously published mitogenomes to compare mitochondrial genome diversity across a large part of the Cape buffalo distribution range and discuss the conservation implications of our findings. With the data presented here (and in previous studies) we aim to ultimately assist in the genetic management and conservation of Cape buffalo populations within, and beyond, South Africa. This is particularly relevant given the importance of genetic assessment and monitoring of wildlife populations now recognised at national and international levels (Hoban et al., 2020; Hoban et al., 2021; Laikre et al., 2020).

Methods

Mitogenome assembly and annotation

Mitogenomes were assembled using NOVOPlasty v4.0 (Dierckxsens, Mardulyn, & Smits, 2017) from the cleaned and trimmed reads of 40 Cape buffalo nuclear genomes sequenced by de Jager et al. (2021). NOVOPlasty assembles organellar genomes from whole genome sequencing data using a seed-and-

extend algorithm (Dierckxsens et al., 2017). The 40 buffalo samples originate from four protected areas in South Africa, namely Addo Elephant National Park (NP) (n = 5), Hluhluwe-iMfolozi Park (n = 15), Kruger NP (n = 15), and Mokala NP (n = 5) – an introduced population established with Kruger buffalo (de Jager et al., 2020). We used the mitogenome of a Cape buffalo from Masai Mara National Reserve in Kenya (GenBank accession: JQ235542 (Heller et al., 2012)) as the seed for the NOVOPlasty assembly. Genome range was set to 15,000-18,000 base pairs (bp) and k-mer set to 33. NOVOPlasty parameter settings were the same for all samples and are provided in a configuration file, an example of which is available here: https://github.com/DeondeJager/Buffalo_Mitogenomics. The assembled mitogenomes were manually checked for ambiguous bases in Geneious Prime v2021.2.2. For the 12 mitogenomes that had one or more ambiguous bases, the cleaned and trimmed genome reads were mapped against the assembled mitogenome using the BWA-MEM algorithm in bwa v0.1.17 (Li & Durbin, 2009). The resultant BAM file was sorted by position using SAMtools v1.9 (Li et al., 2009) and subsequently viewed in Integrative Genomics Viewer (IGV) v2.5.3 (Robinson et al., 2011; Thorvaldsdóttir, Robinson, & Mesirov, 2012) and ambiguities manually resolved where possible. The assembled mitogenomes were *de novo* annotated with MITOS2 (Donath et al., 2019). The MITOS2 annotations were manually curated through comparisons with annotations of the cattle (*Bos taurus*) RefSeq reference mitogenome (NC_006853.1) and the African buffalo reference mitogenome (NC_020617.1) (Hassanin et al. 2012). Specifically, all protein-coding genes were translated in Geneious using the vertebrate mitochondrial codon table (transl_table 2) and the protein lengths were compared to those from the two reference mitogenomes. The mitogenome nucleotide sequence data reported are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK062533-BK062572.

Genetic diversity and structure

To investigate whether there was any sub-structuring of mitogenomes across the range of Cape buffalo, the 40 South African mitogenomes were combined with the 43 from Heller et al. (2012), originating from Ethiopia, Kenya, Zimbabwe, and Botswana, and the mitogenome from Hassanin et al. (2012) from Tanzania, totalling 84 mitogenomes (Table S1). The 13 protein-coding genes, two rRNA genes and D-loop were extracted, concatenated, and aligned using MUSCLE (Edgar, 2004). The alignments were imported into MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) in FASTA format and exported in nexus format, excluding ambiguous sites, for a total alignment length of 13,786 bp. The alignment was used to construct a minimum spanning network (Bandelt, Forster, & Röhl, 1999) of haplotypes in PopArt v1.7 (http://popart.otago.ac.nz) (Leigh & Bryant, 2015). PopArt automatically masks sites with >5% missing data (gaps or ambiguous nucleotides). Diversity and divergence statistics were calculated in DnaSP v6.12.03 (Rozas et al., 2017), ignoring gaps (see Table 1 for number of sites).

The D-loop regions of the 84 Cape buffalo mitogenomes were extracted and combined with publicly available D-loop sequences on GenBank of all African buffalo (*S. caffer*) subspecies from Simonsen et al. (1998), van Hooft et al. (2002), and Smitz et al. (2013), giving a final dataset of 876 sequences (Table S2). Sequences were aligned using MAFFT (Katoh & Standley, 2013) and cleaned and trimmed using Gblocks (Castresana, 2000; Talavera & Castresana, 2007), both implemented on the NGPhylogeny.fr platform (Lemoine et al., 2019). Gblocks parameters were set as follows: Minimum number of sequences for a conserved position (b1) = 50% of sequences + 1, Minimum number of sequences for a flank position (b2)

= 85% of sequences, Maximum number of contiguous non-conserved positions (b3) = 8, Minimum length of a block (b4) = 10, Allowed gap positions (b5) = "With half". The resultant alignment, consisting of 285 aligned bases, was used to construct a minimum spanning network of haplotypes in PopArt.

A Bayesian phylogenetic tree was constructed by aligning whole mitogenomes using MAFFT, with the water buffalo (*Bubalus bubalis*) reference mitogenome used as the outgroup (NC_049568.1), giving an alignment of 16,418 bp. The best substitution model was determined to be HKY+G+I, using the Bayesian information criterion in jModelTest v2.1.10 (Darriba, Taboada, Doallo, & Posada, 2012; Guindon & Gascuel, 2003), with 5 substitution schemes. The tree was constructed in MrBayes v3.2.6 (Ronquist et al. 2012), with two independent runs with four chains and five million generations each, sampling trees every 100 generations, and using a burn-in of 25%. Convergence was reached both within and between the two runs, as indicated by the ESS values >100, the potential scale reduction factor (PSRF) being very close to 1 for all parameters, and the standard deviation of split frequencies being less than the suggested 0.01. The tree was visualised in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and further processed in InkScape v1.1.2 (https://inkscape.org/).

Validation of the divergent mitogenome

The divergent mitogenome in sample A_268_14, a male buffalo from Addo Elephant NP, was apparent from the mitogenomic alignment and haplotype network. To ensure that this mitogenome was not an artefact due to an assembly error, or a nuclear copy of the mitogenome (NuMT), we performed several quality checks. First, it was confirmed that all the protein-coding genes had no internal stop codons and were predicted to encode full-length proteins. Second, the genomic reads were mapped to the assembled mitogenome using the BWA-MEM algorithm in bwa v0.1.17 and the BAM file visualised in IGV to check for any obvious assembly errors, of which none were identified.

Finally, a segment of the cytochrome *b* gene was sequenced using Sanger sequencing, not only of the sample in question, but also of its two closest relatives from Addo Elephant NP: A_05_08 (male) and A_247_14 (female), as determined with 11 microsatellite loci in de Jager, Swarts, Harper, and Bloomer (2017). DNA was extracted from whole blood using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. A 354 bp internal portion of the cytochrome *b* gene was amplified using the primers L14841 and H15149 (Kocher et al., 1989) (L14841: 5'-[AAAAAGCTTCCAT]CCAACATCTCAGCATGATGAAA-3', H15149: 5'-[AAACTGCAG]CCCCTCAGAATGATATTTGTCCTCA-3', where the square brackets indicate extra sequence present in the primers designed by Kocher et al. (1989), but not included in the primers used in this study). PCR amplification was performed in reactions containing 1X Standard Reaction Buffer, 2.5 mM $MgCl_2$, 4 mM dNTP mix, 1X bovine serum albumin, 3 U Super-Therm Taq DNA polymerase (Thermo Fisher Scientific, Massachusetts, United States), 0.1 µM L14841 forward primer, 0.1 µM H15149 reverse primer, 10-30 ng DNA, and molecular grade water to a final reaction volume of 25 µL, with the following conditions in a SimpliAmp Thermal Cycler (Life Technologies, California, United States): 95˚C for 3 min followed by 25 cycles of 94˚C for 30 s, 55˚C for 30 s, 72˚C for 30 s, and a final extension step at 72˚C for 7 min. PCR products were cleaned via ethanol precipitation and 2.5 µL clean product used in a cycle sequencing reaction (BigDye© Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific)) containing 1X Ready

Reaction Premix, 1X BigDye© Sequencing Buffer, 0.5 µM L14841 forward primer or 0.1 µM H15149 primer and molecular grade water to a final volume of 10 µL. Reactions were performed in a SimpliAmp Thermal Cycler with 25 cycles at 96˚C for 10 s, 50˚C for 5 s and 60˚C for 4 min. Products were cleaned via ethanol precipitation and submitted to the DNA Sequencing Facility of the Faculty of Natural and Agricultural Sciences at the University of Pretoria for Sanger sequencing on a 3500xl automated sequencer (Life Technologies, California, United States). All products were sequenced in both the forward and reverse directions to obtain 2X depth of coverage of each base position. The sequences were cleaned and trimmed in Geneious and the consensus sequences were aligned using MUSCLE to the cytochrome *b* genes extracted from the five assembled mitogenomes from Addo Elephant NP. The alignments were trimmed to a final length of 237 bp and used to construct a minimum spanning network of haplotypes in PopArt. The nucleotide sequences generated via Sanger sequencing are available on GenBank under the following accession numbers: OP485291-OP485293.

Divergence dating

To estimate the time of divergence between the divergent mitogenome and other Cape buffalo mitogenomes, and to determine its phylogenetic position within Bovini, we constructed a dated phylogeny in BEAST v2.5.2 (Bouckaert et al., 2019). The 13 protein-coding genes, two rRNA genes (12S and 16S), and the D-loop were extracted, concatenated, and aligned using MAFFT, from the following mitogenomes: the divergent *S. c. caffer* mitogenome from Addo Elephant NP (A_268_14), another *S. c. caffer* from Addo Elephant NP (A_87_13), and a third *S. c. caffer* from Tanzania (GenBank accession: EF536353)), eight other Bovidae mitogenomes included in Hassanin et al. (2012), namely *Tragelaphus oryx* (JN632704), *Bos taurus taurus* (EU177832), *Bos taurus indicus* (EU177868), *Bos grunniens* (NC_006380), *Bison bison* (JN632601), *Bubalus bubalis bubalis* (AF547270), *Bubalus bubalis carabanesis* (JN632607), and the dwarf musk deer (*Moschus berezovskii*, NC_012694) belonging to the Moschidae family, also included in Hassanin et al. (2012), which was used as an outgroup. The alignment was defined as three partitions: the 13 protein coding genes (11,394 sites), the two rRNA genes (2,551 sites), and D-loop (1,027 sites). The best-fit substitution model for each partition was estimated in jModeltest v2.1.10 and determined to be GTR+G for the protein-coding and rRNA partitions, and HKY+G (Hasegawa, Kishino, & Yano, 1985) for the D-loop partition, all with four rate categories, based on the Bayesian information criterion.

The evolutionary model was prepared in BEAUTi v2.6.5 (Bouckaert et al., 2019), with independent (unlinked) site and clock models, a linked tree model, empirical base frequencies, a relaxed log normal clock (Drummond, Ho, Phillips, & Rambaut, 2006) for each partition with a clock rate of 1.0e-6, and a Yule model for the tree prior. Two fossil calibration points from Bibi (2013) were used to estimate divergence times between sequences in the tree. The first was crown Bovidae, which contained all taxa except *M. berezovskii* and was set as a normal distribution with a mean of 18 million years ago (MYA) and a sigma of 1.0, giving a distribution with a 2.5% quantile of 16 MYA and a 97.5% quantile of 20 MYA. The second was crown Bovini, which contained all taxa except *M. berezovskii* and *T. oryx* and was set as a normal distribution with a mean of 9 MYA and a sigma of 1.0, giving a distribution with a 2.5% quantile of 7.04 MYA and a 97.5% quantile of 11.0 MYA. Monophyly was enforced in both cases. The analysis was executed in two independent runs each with an MCMC chain length of 20,000,000 generations, sampling every 1,000 runs and a burn-in of 2,000 (10%). The xml file (available at

https://github.com/DeondeJager/Buffalo_Mitogenomics) was executed in BEAST and the resulting log files were examined in Tracer v1.7.2 (Rambaut, Drummond, Xie, Baele & Suchard, 2018), which showed that all parameters converged, with individual and combined effective sample sizes greater than 200, and in most cases greater than 500. The tree files from both runs were combined using LogCombiner (Bouckaert et al., 2019) and the output was used in Tree Annotator (Bouckaert et al., 2019) to build a maximum clade credibility (MCC) tree with a 20% burn-in (4,000 trees), a posterior probability limit of 0.5, and using median node heights. The MCC tree was visualised in FigTree and further processed in InkScape.

Results and discussion

South African mitogenomes

The 40 South African (SA) Cape buffalo mitogenomes were each assembled into a single, circularised contig using NOVOPlasty. Coverage ranged from 154X to 1,036X (mean: 419X) and lengths were 16,357-16,362 bp (Table S3), which was similar to the sizes of published mitogenomes for the subspecies (16,357-16,361 bp) (Hassanin et al., 2012; Heller et al., 2012). The cause of these slight size differences was usually one or two base pair indels in the rRNA genes or D-loop region. There were two exceptions to this, with a mitogenome from Kruger NP (B98_597) having two insertions of 5 bp and 3 bp in the D-loop, increasing its size to 16,367 bp, and a mitogenome from Addo Elephant NP (A_268_14) having a 10 bp insertion in the D-loop increasing its size to 16,375 bp. A large deletion of 74 bp in the D-loop of a Cape buffalo mitogenome from Ethiopia was previously reported by Heller et al. (2012), showing that large indels can be present in the D-loop within the subspecies.

The MITOS2 annotations were mostly accurate, as the protein-coding genes of all 40 assembled SA mitogenomes had no internal stop codons and were predicted to produce proteins of the expected length, as compared to the cattle and African buffalo reference mitogenomes. The exceptions were the genes *cox3* and *nad5*, which were predicted to produce proteins that were respectively one and three amino acids longer than the references in all 40 samples. This was due to the 3' end of *cox3* predicted by MITOS2 to be four base pairs downstream, and the start codon of *nad5* predicted to be nine base pairs upstream, compared to the references. These annotations were manually corrected to match the annotations of the reference mitogenomes and produce proteins of the expected lengths. Two samples from Addo Elephant NP, A_243_14 and A_251_14, had an A to G transition in the second position of the stop codon of the *atp8* gene (*m.8326A>G*). This change was supported by 500/536 reads (93%) at this position in sample A_243_14 containing a G, with only 35 reads (7%) containing an A and one read (0%) containing a C. For sample A_251_14, 347/376 reads (92%) contained a G at this position, with only 27 reads (7%) containing an A and two reads (1%) containing a C. The substitution changes the stop codon (UAA) to one coding for the amino acid tryptophan (UGA). However, there is an in-frame stop codon six base pairs downstream (UAG), resulting in the predicted protein product being 68 amino acids long, as opposed to the expected 66 amino acids, with a tryptophan (UGA) and serine (UCC) added to the C-terminal of the ATP8 protein. All the above results indicated that the mitogenomes assembled from the whole genome sequencing reads were of high quality and accuracy.

Genetic diversity and structure

The Addo Elephant NP buffalo had the highest nucleotide diversity of all Cape buffalo populations in this study, almost three times that of the next most diverse population (Chobe NP) (Table 1). This was surprising, as the buffalo population from Addo Elephant NP is known to have the lowest nuclear genetic diversity of all natural South African populations, due to a strong historical bottleneck and subsequent isolation (de Jager et al., 2021; de Jager et al., 2020; O'Ryan et al., 1998). This high diversity was driven by one sample, A_268_14, the mitogenome of which was highly divergent, not only from the other Addo Elephant NP mitogenomes, but also those of the rest of the subspecies. Interestingly, the nuclear genome of this individual did not show any unique signals in analyses conducted by de Jager et al. (2021) and clustered with the nuclear genomes of the other Addo Elephant NP buffalo. The average number of nucleotide substitutions between this mitogenome and the rest of Cape buffalo (n = 83) was 0.03517, a divergence of 3.5%. When this sample was excluded, Addo Elephant NP buffalo had the lowest nucleotide diversity, five times lower than that of the next lowest population (Hluhluwe-iMfolozi Park) (Table 1), which is more in line with the nuclear diversity results from previous studies and the population's demographic history. Hluhluwe-iMfolozi Park, having also experienced an historical bottleneck, though seemingly not as extreme as that of Addo Elephant NP (de Jager et al., 2021; O'Ryan et al., 1998), harbours the second largest free-ranging buffalo population in South Africa (~4,544 individuals in 2019 (EKZNW, 2019)), but had the lowest haplotype diversity with only four haplotypes from 15 samples (Table 1). Additionally, its nucleotide diversity was second lowest and substantially lower than that of Mokala NP and Kruger NP, the latter of which harbours the largest free-ranging buffalo population in South Africa (~40,900 individuals in 2011 (Cornélis et al., 2014)).

Nucleotide diversity was higher in southern Africa (Zimbabwe, Botswana, SA) than in East Africa (Kenya, Ethiopia), even when excluding the divergent mitogenome from the Addo Elephant NP buffalo, whereas higher haplotype diversity was observed in East Africa (Table 1). It is hypothesised, based on genetic and fossil evidence, that Cape buffalo colonised southern Africa from East Africa during the last ~80,000 years (Heller et al., 2012; Smitz et al., 2013). While our nucleotide diversity results may not seem to support this scenario (where diversity is expected to be higher in the source compared to the sink population), we note that there is only one mitogenome available from Tanzania and none from Uganda. The former has the largest free-ranging Cape buffalo population, upwards of 100,000 individuals (Cornélis et al., 2014), while the region of present-day Uganda appears to harbour several ancestral mitochondrial lineages and unique haplotypes of Cape buffalo (Smitz et al., 2013). Inclusion of mitogenomes from these two locations would more accurately represent the genetic diversity of East African buffalo and likely reflect the expected source-sink pattern. Furthermore, genetic diversity and migration rates are high throughout most of the range of the subspecies (Simonsen et al., 1998; Smitz et al., 2013; van Hooft et al., 2000), making other evolutionary patterns more difficult to discern, while recent demographic events (e.g. bottlenecks in the last ~300 years, exacerbated by subsequent isolation) of individual populations may mask signals of previous evolutionary or demographic events (Excoffier & Schneider, 1999; Rasmus Heller et al., 2012).

Table 1 Summary statistics of mitogenome alignments of various subsets of the data calculated in DnaSP v6.12.03.

Dataset	Sample size (N)	Segregating sites (S)	No. of haplotypes (h)	Haplotype diversity (Hd)	Nucleotide diversity (π)	Average no. of nucleotide differences (k)	Tajima's *D*	No. of sites in alignment (bp)
Addo Elephant National Park	5	518	3	0.8	0.01526	209.2	-1.21246	13,706
Addo Elephant National Park (excl. divergent)	4	10	2	0.667	0.00049	6.667	2.22343	13,710
Kruger National Park	15	228	12	0.971	0.00474	64.971	-0.34306	13,707
Mokala National Park	5	152	5	1	0.00481	65.9	-0.78218	13,707
Hluhluwe-iMfolozi Park	15	95	4	0.467	0.00258	35.41	0.92808	13,705
South Africa	**40**	**702**	**23**	**0.919**	**0.00626**	**85.813**	**-1.81535***	**13,702**
South Africa (excl. divergent)	**39**	**287**	**22**	**0.915**	**0.00465**	**63.682**	**-0.24381**	**13,704**
Botswana (Chobe NP)	8	203	7	0.964	0.00537	73.607	-0.32758	13,706
Zimbabwe (Hwange NP)	8	149	5	0.857	0.00451	61.786	0.37156	13,705
Kenya (Masai Mara NR)	20	228	14	0.953	0.00372	50.984	-0.88809	13,704
Ethiopia (Omo Valley)	7	134	7	1	0.00342	46.667	-0.89648	13,633
Heller et al. 2012 (recalculation)	**43**	**389**	**30**	**0.978**	**0.00439**	**59.886**	**-1.25622**	**13,627**
Heller et al. 2012 (original)	**43**	**456**	**30**	**-**	**0.00435**	**-**	**-1.24**	

Dataset	Sample size (N)	Segregating sites (S)	No. of haplotypes (h)	Haplotype diversity (Hd)	Nucleotide diversity (π)	Average no. of nucleotide differences (k)	Tajima's *D*	No. of sites in alignment (bp)
Southern Africa (South Africa, Botswana, Zimbabwe)	**56**	**780**	**33**	**0.953**	**0.00604**	**82.764**	**-1.85935***	**13,699**
Southern Africa (excl. divergent)	**55**	**382**	**32**	**0.951**	**0.0049**	**67.141**	**-0.71192**	**13,701**
East Africa (Tanzania, Kenya, Ethiopia)	**28**	**277**	**21**	**0.974**	**0.00362**	**49.323**	**-1.23484**	**13,629**
All mitogenomes	**84**	**857**	**53**	**0.976**	**0.00517**	**70.476**	**-2.04442***	**13,624**
All mitogenomes (excl. divergent)	**83**	**488**	**52**	**0.975**	**0.00444**	**60.51**	**-1.32907**	**13,625**

*$P < 0.05$

The minimum spanning network of mitogenomes showed no clear evidence of haplotypes clustering according to location and reflected the high diversity within the subspecies, as indicated by few shared haplotypes and many mutational steps between haplotypes (Figure 1A). There was a tendency for haplotypes from southern Africa to be at the tips of the network, with those from East Africa generally more internal (Figure 1A), which adds support to the east-south colonisation hypothesis (Smitz et al., 2013). The Bayesian phylogenetic tree also showed no clear relationship between mitogenomes and geographical origin at a broad scale (i.e., between eastern and southern Africa), which was well supported with most nodes having a posterior probability of one or close to one (Figure S1).

A highly divergent mitogenome in Addo Elephant NP

The divergent mitogenome from an Addo Elephant NP buffalo encoded full length proteins with no internal stop codons, thus was likely not an assembly error or a NUMT. This mitogenome did not cluster with the others from Addo Elephant NP in the haplotype network (Figure 1A) or the phylogenetic tree (Figure S1). The closest haplotypes were the most common haplotype in Kenya and another from Zimbabwe, although there were 472 substitutions between each of these and the divergent mitogenome. To further validate the authenticity of the divergent mitogenome, we performed Sanger sequencing of a fragment of the cytochrome *b* gene of this sample (A_268_14, male), as well as its two closest relatives, A_05_08 (male, Wang pairwise relatedness estimate (r) = 0.45) and A_247_14 (female, r = 0.31), and constructed a haplotype network with the cytochrome *b* genes of other assembled Addo mitogenomes (Figure 1B). We found that the Sanger-generated sequence of A_268_14 was identical to that of the assembled sequence, indicating again that it was unlikely to be an assembly error or NUMT. Furthermore, its closest relative (A_05_08) had an identical Sanger-generated sequence and thus shares this divergent mitogenome with sample A_268_14. Two other haplotypes were identified, separated from the divergent haplotype by 13 and 14 mutations across 237 bp, which were shared among the five other Addo samples, including the next-closest relative of A_268_14, A_247_14 (Figure 1B).

The divergent mitogenome might have a relatively high frequency in the Addo population, as it was present in one of five (20%) unrelated individuals selected for whole genome sequencing in de Jager et al. (2021). This frequency increases to two in seven (28.5%) if the Sanger sequencing results are included, though this sampling was biased towards close relatives of A_268_14. Nevertheless, the sample size was relatively low compared to the census size of Addo (~800 buffalo) and thus likely does not adequately represent the frequency of the divergent haplotype in the population. However, it is evidently at a high enough frequency to be detected from just five samples. The Addo population experienced a relatively severe bottleneck in the late 1800s and early 1900s, and remained at low numbers (<250) throughout much of the 1900s, with the lowest known census size being 52 individuals in 1985 (*Pers. Comm.* D. Zimmerman 2015; (de Jager et al., 2021)), resulting in the loss of genetic diversity through drift and inbreeding (de Jager et al., 2021; de Jager et al., 2020; O'Ryan et al., 1998). Despite this, the divergent mitochondrial lineage was maintained in the population, pointing towards a high pre-bottleneck frequency. Addo buffalo have been used to seed and supplement other buffalo populations in southern Africa and have been sold to private wildlife ranchers for many years, raising the possibility that the divergent lineage may also now be present in those populations and ranches. Future studies could robustly estimate the frequency of this haplotype in the Addo population through more extensive

sampling and investigate whether there are any potential fitness benefits for individuals that carry this divergent lineage.

To determine where the divergent mitogenome clusters within the diversity of the entire species, we constructed a minimum spanning haplotype network of 285 bp of the D-loop sequence from all African buffalo subspecies (Figure 1C). We were able to reconstruct lineages H1 and H2 as identified in Smitz et al. (2013), where H1 is the South-Eastern lineage (mainly Cape buffalo, *S. c. caffer*) and H2 is the West-Central lineage (predominantly containing the other three subspecies, *S. c. nanus*, *S. c. aequinoctialis*, *S. c. brachyceros*). The lineages are not monophyletic, and each contains some discordant haplotypes (Figure 1C, also see Figure 1 in Smitz et al. (2013)). The D-loop haplotype of the divergent Addo sequence (of *S. c. caffer* origin) grouped with the H2 (West-Central) lineage and not the H1 (South-Eastern) lineage like most other *S. c. caffer* sequences (Figure 1C). It was at a tip of the network, separated by five mutational steps from a relatively common haplotype from Western Africa that was also present in West Central, and Central Africa (Figure 1C). Its location in the H2 lineage, and the number of mutations separating it from its nearest haplotype, is not especially unique in this network. There are several haplotype-lineage "mismatches" and other haplotypes that are separated by four or six mutations, the latter always involving haplotypes from Uganda (Figure 1C). This is interesting, because it highlights the limited power of using only a portion of the D-loop, as high diversity in this region might be expected and highly divergent lineages such as the one in Addo may go undetected (or under-investigated). Including a mitochondrial gene (or fragment thereof) would have unequivocally identified this lineage as highly divergent and unique. It also raises the question of whether the other relatively diverged haplotypes in Figure 1C are as highly divergent across the rest of the mitogenome as the Addo mitogenome is, or whether the variation is localised to the D-loop.

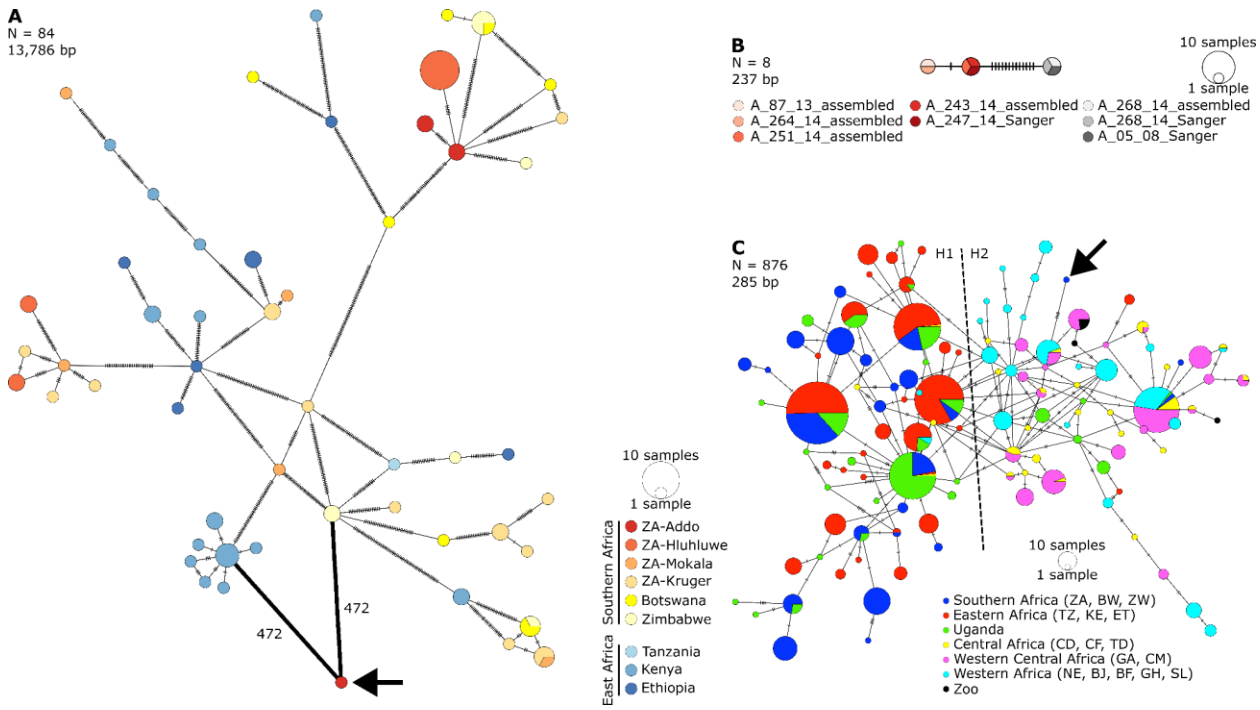

Figure 1 Minimum spanning haplotype networks of: **A** the 13,786 bp alignment of 13 protein-coding genes, rRNA genes, and D-loop from 44 publicly available and 40 newly generated Cape buffalo mitogenomes, **B** the 237 bp alignment of cytochrome *b* of Addo samples from both assembled mitogenomes and those generated via Sanger sequencing, and **C** the 285 bp alignment of publicly available D-loop sequences of all African buffalo subspecies and from the 40 newly generated mitogenomes in this study. Each line on a link connecting haplotypes represents one substitution. The arrows in **A** and **C** indicate the divergent mitogenome from Addo (sample A_268_14). The numbers in **A** show the number of mutational steps between the divergent mitogenome and the nearest haplotypes. ISO 3166 two-letter country codes used in **A** and **C** refer as follows: ZA – South Africa, BW – Botswana, ZW – Zimbabwe, TZ – Tanzania, KE – Kenya, ET – Ethiopia, CD – The Democratic Republic of the Congo, CF – Central African Republic, TD – Chad, GA – Gabon, CM – Cameroon, NE – Niger, BJ – Benin, BF – Burkina Faso, GH – Ghana, SL – Sierra Leone. "Zoo" refers to samples from Antwerpen Zoo (Belgium, n = 2), Berlin Zoo (Germany, n = 1), Dresden Zoo (Germany, n = 1), and Safari Park Beekse Bergen (Netherlands, n = 1) (van Hooft et al., 2002). All networks were generated in PopArt and further processed in InkScape.

Finally, using a fossil-calibrated phylogeny, we found that the divergent Addo mitogenome formed a sister group to other Cape buffalo, with a divergence date of approximately 2.15 million years ago (Ma) (95% highest posterior density: 1.58-2.76 Ma) (Figure 2). This estimate overlaps with the divergence time of both species-level divergence [bison (*Bison bison*) and yak (*Bos grunniens*)], and subspecies-level divergence (water buffalo subspecies: *Bubalus bubalis bubalis* and *B. b. carabanesis*) within Bovini (Figure 2), an observation which also holds within the broader Bovidae group (Bibi, 2013). All divergence dates of known species-level relationships estimated here (Figure 2) correspond to accepted, published estimates based on mitogenomes (Bibi, 2013) and whole nuclear genomes (Chen et al., 2019), indicating that the divergence date of the Addo mitogenome of interest is accurate.

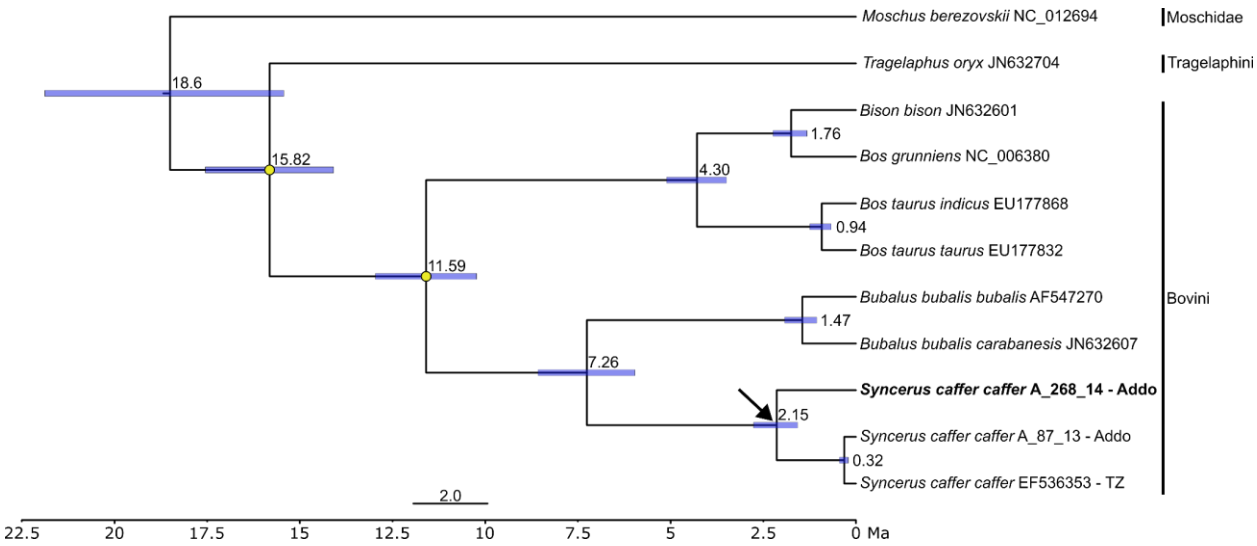

Figure 2 Fossil-calibrated, Bayesian phylogenetic tree of mitogenomes estimated in BEAST. Yellow circles show the fossil calibration points used. Numbers on the nodes show the estimated age (height) of the node in millions of years. Node bars represent the 95% highest posterior density of the node height. The divergent Addo mitogenome is highlighted in bold text, with the arrow indicating the node where it diverged from other Cape buffalo mitogenomes. TZ – Tanzania, Ma – million years ago.

Source of the divergent Addo mitogenome

Given the high level of divergence (3.5% and 2.15 Ma) of this mitogenome from other Cape buffalo mitogenomes, we hypothesise that the most likely source is through introgression with a closely related species or subspecies. In the phylogenetic analysis, the divergent mitogenome unequivocally grouped with Cape buffalo (node posterior support = 1), which excludes water buffalo (*Bubalus bubalis*) and cattle (*Bos taurus*) as potential sources. Importantly, we note that there have been no recorded human-mediated introductions of African buffalo into the Addo population since it was proclaimed as a protected area in 1931 (*Pers. Comm.* D. Zimmerman 2015; (de Jager et al., 2021)), and it therefore represents the long-term, naturally-occurring population and lineages of Cape buffalo in that region of South Africa. Thus, any introgression that may have occurred is assumed to have happened before 1931 and from a naturally co-occurring species, as wildlife translocations were infrequent in South Africa before this time due to the lack of appropriate game capture techniques and technologies (Carruthers, 2008).

A potential source could be introgression from other extant African buffalo subspecies. Unfortunately, no mitogenomes or mitochondrial gene sequences from any of the other three subspecies were publicly available to test this possibility, and while the D-loop haplotype network may lend some support to this, it is based on very limited data and thus may be misleading. Furthermore, the recent split between the two African buffalo lineages (H1: *S. c. caffer* and H2: *S. c. nanus*, *S. c. aequinoctialis*, *S. c. brachyceros*) of between 145,000 and 449,000 years ago (Smitz et al., 2013) and the fact that none of these subspecies co-occur (or have co-occurred) with Cape buffalo in South Africa, indicates that another extant subspecies is unlikely to be the source of the divergent mitogenome. There are currently no other bovines that occur in or around Addo Elephant NP that may be the source of this mitogenome through introgression.

Another possibility may be that this divergent lineage represents retained ancestral polymorphism of an early colonisation of southern Africa by buffalo from eastern or central Africa. The present-day southern African populations are likely the result of a population expansion event from eastern to southern Africa between 80,000 and 50,000 years ago (Heller et al., 2012; Smitz et al., 2013; van Hooft et al., 2002). However, there is radiometric evidence from Kathu Pan in the Northern Cape Province of South Africa that fossils from this site are around 542,000 years old, which includes *Syncerus caffer*, *Syncerus* (*Pelorovis*) *antiquus*, and other undetermined buffalo specimens (Porat et al., 2010). Consequently, Smitz et al. (2013) proposed the possibility that southern Africa was colonised and re-colonised multiple times throughout the Mid- to Late Pleistocene. It may thus be that the divergent mitogenome could be from a remnant population of buffalo from one of these previous colonisation events that survived around present-day Addo and was assimilated into the contemporary population during subsequent colonisation events. Mitogenomic data from other extant subspecies, which were ancestral to the Cape subspecies, may also shed light on this potential scenario.

However, we propose that the most likely source of the divergent mitogenome in the modern Addo buffalo population is the extinct long-horned ("giant") buffalo, *Syncerus antiquus*. This species was widely distributed from southern, to eastern and northern Africa (Klein, 1994) and appears to have gone extinct in southern and eastern Africa at the end of the Pleistocene (~11,700 years ago) (Klein, 1984; Klein, 1994), and in northern Africa around 4,000 years ago (Gautier & Muzzolini, 1991). Cape buffalo are (and were) only distributed south of the Sahara Desert (du Toit, 2005), but the two species are known to have co-occurred in space and time, as shown by fossil evidence from palaeontological sites in South Africa, near Addo Elephant NP, namely Klasies River Mouth Caves, Nelson Bay Cave, and Byneskranskop Cave 1 (Klein, 1994).

The long-horned buffalo has undergone several taxonomic changes (see Faith (2014) for an overview); some authors suggested a separate genus *Pelorovis antiquus* (Gentry & Gentry, 1978), but the growing consensus is for the species to be included under *Syncerus* (as *S. antiquus*) (Faith, 2014). Peters, Gautier, Brink, & Haenen (1994) went as far as to classify the long-horned buffalo as a subspecies of African buffalo, *Syncerus caffer antiquus*, based on the similarity of its postcranial bones to Cape buffalo (*S. c. caffer*). The authors suggested that the *antiquus* buffalo is not an extinct species, but was a long-horned morph of Cape buffalo and represents an example of phyletic change (Peters et al., 1994). However, Klein (1994) disputes this subspecific classification based on horn core morphology and differences in the dental anatomy between *caffer* and *antiquus*, pointing to the fact that many closely related bovid species (accepted as distinct species) have very similar postcranial skeletons and often can only be confidently distinguished based on horn cores and/or dental morphology. Additionally, Klein (1994) notes that no intermediate (hybrid) forms have been found in the fossil record. However, this may be due to the close resemblance of their postcranial anatomy making it difficult to identify hybrids, especially if back-crossing occurs with either parent, or because fossilisation and hybridisation are both rare events, making the preservation and discovery of hybrid fossils likely an ultra-rare event.

Whatever the taxonomic status of the long-horned buffalo, it is evident that it is closely related to African buffalo, and specifically the Cape buffalo subspecies (Klein, 1994; Peters et al., 1994). Intra- and

interspecific hybridisation and introgression have been well-documented in nature across many taxa (Taylor & Larson, 2019), including bovids (Fadakar et al., 2020; Grobler, van Wyk, Dalton, van Vuuren, & Kotze, 2018; Hassanin et al., 2012; Rakotoarivelo, O'Donoghue, Bruford, & Moodley, 2019; van Wyk et al., 2019), supporting the hypothesis that the source of the divergent mitogenome may be due to introgressive hybridisation between Cape and long-horned buffalo. There is no clear reason, to our knowledge, why this introgressive hybridisation process (if that is indeed what has occurred) should have been localised to the southern tip of Africa, as Cape and long-horned buffalo co-occurred in space and time across a much larger part of Africa (Faith, 2014). It could be that it has simply not been detected yet, due to larger population sizes in other regions compared to Addo Elephant NP, which would require a more intensive sampling effort to detect a rare lineage, whereas in Addo the lineage is at a relatively high frequency.

To test the hypotheses of the source of the divergent mitogenome proposed above it is necessary to obtain genetic data from the extinct long-horned buffalo, ancient Cape buffalo from southern South Africa, as well as whole mitogenomes from the other extant African buffalo subspecies. With these data it would be possible to (i) determine the taxonomic position of the extinct long-horned buffalo with respect to extant bovines, (ii) determine whether the divergent mitogenome in Addo buffalo clusters with the long-horned buffalo, and therefore whether it is indeed the result of introgression from this species, (iii) test the phyletic change hypothesis of Peters et al. (1994) (with both ancient Cape buffalo and long-horned buffalo genetic data), and (iv) determine whether the divergent Addo mitogenome clusters with any of the other extant buffalo subspecies, which may support the retained ancestral lineage hypothesis. With the techniques in the fields of ancient DNA and paleogenomics constantly improving, it is now possible to obtain these ancient genetic data, as shown by Hempel et al. (2022) who recovered the nuclear genome of the extinct blue antelope (*Hippotragus leucophaeus*) from a specimen dating to 9,800-9,300 years old from Nelson Bay Cave in South Africa.

Conservation consequences for the Addo Elephant NP buffalo population

Regardless of the source of the divergent mitogenome, the fact that it is found in the present-day Addo buffalo further increases the conservation value of this population and indicates that its genetic management requires careful consideration. This population already has high conservation value in South Africa for several reasons: it is one of three remnant populations that survived the hunting onslaught and rinderpest epidemic in the 1800s and 1900s (the others being Kruger National Park and Hluhluwe-iMfolozi Park); it is the only naturally occurring disease-free population in the country (another disease-free population of Kruger National Park buffalo was established in Mokala National Park and Graspan); and finally, Addo buffalo are used to seed or supplement populations throughout southern Africa and are sold to private wildlife ranchers (de Jager et al., 2020; Laubscher & Hoffman, 2012). However, it has long been suspected and repeatedly shown that this population has low genetic diversity due to the previously-mentioned population bottleneck and subsequent isolation, and that it is threatened by inbreeding (de Jager et al., 2021; de Jager et al., 2020; O'Ryan et al., 1998).

To prevent further inbreeding and inbreeding depression (the manifestation of inbreeding in the biology of the population or individuals, such as reduced population growth, increased susceptibility to disease,

low sperm quality, low body condition, etc.), O'Ryan et al. (1998) calculated that one breeding bull (male) of Kruger stock should be introduced to the Addo population every generation (~7.5 years), which can now be done via the disease-free population in Mokala National Park (de Jager et al., 2020). This frequency of introduction was proposed to prevent genetic swamping of the Addo buffalo with Kruger genetic variation. Wisely, it was also suggested that any introduced individuals be males (O'Ryan et al., 1998). If this guidance is followed, any unique mitochondrial diversity, such as the highly divergent lineage identified here, should be retained in the Addo population (if any female buffalo carry this lineage), because mitochondrial DNA is inherited only from mother to offspring.

Conclusion

The findings in this study highlight the importance of conserving distinct genetic units within a species or subspecies and the value of a cautionary approach to genetic management of populations. Previous genetic studies based on nuclear DNA found that while Addo, Hluhluwe-iMfolozi and Kruger buffalo populations are separate genetic clusters, it did not seem that there was much unique diversity within Addo, and that it contained a subset of the genetic variation found in Kruger (de Jager et al., 2021; de Jager et al., 2020; O'Ryan et al., 1998). This is true to a certain extent when looking at nuclear DNA, but this study has shown that all three South African populations contain unique mitochondrial lineages, and that the Addo population contains this highly divergent, cryptic mitochondrial lineage that has not yet been detected in any other Cape or African buffalo populations. Whether this mitochondrial lineage confers any fitness or adaptive advantage to Addo buffalo is not yet known, but it has persisted in the population for potentially thousands of years alongside other, more common, Cape buffalo lineages, and this further increases the conservation value of this vulnerable population. If the Addo population were allowed to go extinct, or swamped with Kruger buffalo, this lineage would probably have been lost. Thus, an intermediate position, where genetic clusters are preserved as far as possible, but inbreeding depression is prevented, emerges as a sensible approach to genetic management of these populations.

Supplementary material

Supplementary Table S1 Metadata of mitogenomes used in this study.

Supplementary Table S2 Metadata of D-loop sequences used in this study.

Supplementary Table S3 Assembly statistics of the mitogenomes assembled in this study with NOVOPlasty v4.0 from the reads of whole nuclear genomes sequenced by de Jager et al. (2021).

Supplementary Figure S1 Bayesian phylogenetic tree of whole mitogenomes (16,418 bp alignment) constructed in MrBayes, with proportional branch lengths (set in FigTree) for visual clarity. The water buffalo reference mitogenome was used as the outgroup. Branches are coloured by posterior probability, as indicated by the colour scale bar. Node labels indicate nodes that have a posterior probability of less than one. Thus, all unlabelled nodes have a posterior probability of one. Sample names are coloured by country of origin, and by protected area of origin in the case of South African samples, as in Figure 1A. Sample names are followed by the ISO 3166 two-letter code of the country of origin, which refer as

follows: ZA – South Africa, BW – Botswana, ZW – Zimbabwe, TZ – Tanzania, KE – Kenya, ET – Ethiopia. Those samples from South Africa (ZA) are also followed by the protected area of origin.

Acknowledgements

The Centre for High Performance Computing (CHPC), Cape Town, South Africa, is hereby acknowledged for providing the computing resources used to conduct the analyses in this study. The DNA Sequencing Facility of the University of Pretoria is acknowledged. We further acknowledge the support of the DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. DdJ acknowledges postdoctoral research funding from the University of Pretoria and the National Research Foundation (grant ID: 111962). Opinions, findings and conclusions, or recommendations expressed herein are those of the researchers and do not necessarily reflect those of the funding bodies.

References

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*(1), 37-48.

Bibi, F. (2013). A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolutionary Biology, 13*, 166-181. doi:10.1186/1471-2148-13-166

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., . . . Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 15*(4), e1006650. doi:10.1371/journal.pcbi.1006650

Carruthers, J. (2008). "Wilding the farm or farming the wild"? The evolution of scientific game ranching in South Africa from the 1960s to the present. *Transactions of the Royal Society of South Africa, 63*(2), 160-181. Retrieved from http://reference.sabinet.co.za/webx/access/electronic_journals/royalsa/royalsa_v63_n2_a5.pdf

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution, 17*(4), 540-552. doi:10.1093/oxfordjournals.molbev.a026334

Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., . . . Wang, W. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science, 364*(6446). doi:10.1126/science.aav6202

Cornélis, D., Melletti, M., Korte, L., Ryan, S. J., Mirabile, M., Prin, T., & Prins, H. H. T. (2014). African buffalo (Syncerus caffer Sparrman, 1779). In M. Melletti & J. Burton (Eds.), *Ecology, Evolution and Behaviour of Wild Cattle: Implications for Conservation* (pp. 326-372). Cambridge, UK: Cambridge University Press.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772-772. doi:10.1038/nmeth.2109

de Jager, D., Glanzmann, B., Möller, M., Hoal, E., van Helden, P., Harper, C., & Bloomer, P. (2021). High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Scientific Reports, 11*(1), 4540. doi:10.1038/s41598-021-83823-8

de Jager, D., Harper, C. K., & Bloomer, P. (2020). Genetic diversity, relatedness and inbreeding of ranched and fragmented Cape buffalo populations in southern Africa. *PLoS One, 15*(8), e0236717. doi:10.1371/journal.pone.0236717

de Jager, D., Swarts, P., Harper, C., & Bloomer, P. (2017). Friends and family: A software program for identification of unrelated individuals from molecular marker data. *Molecular Ecology Resources, 17*(6), e225-e233. doi:10.1111/1755-0998.12691

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research, 45*(4), e18-e18. doi:10.1093/nar/gkw955

Donath, A., Jühling, F., Al-Arab, M., Bernhart, S. H., Reinhardt, F., Stadler, P. F., . . . Bernt, M. (2019). Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research, 47*(20), 10543-10552. doi:10.1093/nar/gkz833

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology, 4*(5), e88. doi:10.1371/journal.pbio.0040088

du Toit, J. T. (2005). Order RUMINANTIA. In C. T. Chimimba & J. D. Skinner (Eds.), *The Mammals of the Southern African Sub-region* (3rd ed., pp. 616-714). Cambridge: Cambridge University Press.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

EKZNW. (2019). *Annual Report*. Retrieved from http://www.kznwildlife.com/Documents/AnnualReport/EzemveloAR2019.pdf

Excoffier, L., & Schneider, S. (1999). Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proceedings of the National Academy of Sciences, 96*(19), 10597-10602. doi:10.1073/pnas.96.19.10597

Fadakar, D., Malekian, M., Hemami, M. R., Lerp, H., Rezaei, H. R., & Bärmann, E. V. (2020). Repeated hybridization of two closely related gazelle species (Gazella bennettii and Gazella subgutturosa) in central Iran. *Ecology and Evolution, n/a*(n/a). doi:10.1002/ece3.6774

Faith, J. T. (2014). Late Pleistocene and Holocene mammal extinctions on continental Africa. *Earth-Science Reviews, 128*, 105-121. doi:10.1016/j.earscirev.2013.10.009

Gautier, A., & Muzzolini, A. (1991). The life and times of the giant buffalo alias *Bubalus/Homoioceras/Pelorovis antiquus* in North Africa. *Archaeozoologia, 4*, 39–92.

Gentry, A. W., & Gentry, A. (1978). Fossil Bovidae (Mammalia) of Olduvai Gorge, Tanzania; Part I. *Bulletin of the British Museum Natural History Geology Series, 29*, 289-446. Retrieved from https://www.biodiversitylibrary.org/item/112788#page/315/mode/1up

Grobler, P., van Wyk, A. M., Dalton, D. L., van Vuuren, B. J., & Kotze, A. (2018). Assessing introgressive hybridization between blue wildebeest (*Connochaetes taurinus*) and black wildebeest (*Connochaetes gnou*) from South Africa. *Conservation Genetics, 19*(4), 981-993. doi:10.1007/s10592-018-1071-x

Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology, 52*(5), 696-704. doi:10.1080/10635150390235520

Hasegawa, M., Kishino, H., & Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution, 22*(2), 160-174. doi:10.1007/BF02101694

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B. J., Matthee, C., . . . Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies, 335*(1), 32-50. doi:10.1016/j.crvi.2011.11.002

Heller, R., Brüniche-Olsen, A., & Siegismund, H. R. (2012). Cape buffalo mitogenomics reveals a Holocene shift in the African human–megafauna dynamics. *Molecular Ecology, 21*(16), 3947-3959. doi:10.1111/j.1365-294X.2012.05671.x

Heller, R., Lorenzen, E. D., Okello, J. B. A., Masembe, C., & Siegismund, H. R. (2008). Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Molecular Ecology, 17*(22), 4845-4858. doi:10.1111/j.1365-294X.2008.03961.x

Heller, R., Okello, J. B. A., & Siegismund, H. (2010). Can small wildlife conservancies maintain genetically stable populations of large mammals? Evidence for increased genetic drift in geographically restricted populations of Cape buffalo in East Africa. *Molecular Ecology, 19*(7), 1324-1334. doi:10.1111/j.1365-294X.2010.04589.x

Hempel, E., Bibi, F., Faith, J. T., Koepfli, K.-P., Klittich, A. M., Duchêne, D. A., . . . Westbury, M. V. (2022). Blue turns to grey - Palaeogenomic insights into the evolutionary history and extinction of the blue antelope (Hippotragus leucophaeus). *bioRxiv*, 2022.2004.2012.487785. doi:10.1101/2022.04.12.487785

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., . . . Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation, 248*, 108654. doi:10.1016/j.biocon.2020.108654

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., . . . Vernesi, C. (2021). Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. *Bioscience*. doi:10.1093/biosci/biab054

IUCN. (2019). IUCN SSC Antelope Specialist Group. 2019. Syncerus caffer. The IUCN Red List of Threatened Species 2019: e.T21251A50195031. doi:10.2305/IUCN.UK.2019-1.RLTS.T21251A50195031.en. Downloaded on 23 July 2020

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution, 30*(4), 772-780. doi:10.1093/molbev/mst010

Klein, R. G. (1984). Mammalian extinctions and the Stone Age people in Africa. In P. S. Martin & R. G. Klein (Eds.), *Quaternary Extinctions: A prehistoric Revolution* (pp. 553-573). Tucson, Arizona: University of Arizona Press.

Klein, R. G. (1994). The long-horned African buffalo (*Pelorovis antiquus*) is an extinct species. *Journal of Archaeological Science, 21*(6), 725-733. doi:10.1006/jasc.1994.1072

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, 86*(16), 6196-6200. doi:10.1073/pnas.86.16.6196

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

Laikre, L., Hoban, S., Bruford, M. W., Segelbacher, G., Allendorf, F. W., Gajardo, G., . . . Vernesi, C. (2020). Post-2020 goals overlook genetic diversity. *Science, 367*(6482), 1083-1085. doi:10.1126/science.abb2748

Laubscher, L., & Hoffman, L. (2012). An overview of disease-free buffalo breeding projects with reference to the different systems used in South Africa. *Sustainability, 4*(12), 3124-3140. doi:10.3390/su4113124

Leigh, J. W., & Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution, 6*(9), 1110-1116. doi:10.1111/2041-210X.12410

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research, 47*(W1), W260-W265. doi:10.1093/nar/gkz303

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

O'Ryan, C., Harley, E. H., Bruford, M. W., Beaumont, M., Wayne, R. K., & Cherry, M. I. (1998). Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation, 1*(2), 85-94. doi:10.1111/j.1469-1795.1998.tb00015.x

Peters, J., Gautier, A., Brink, J. S., & Haenen, W. (1994). Late quaternary extinction of ungulates in sub-Saharan Africa: A reductionist's approach. *Journal of Archaeological Science, 21*(1), 17-28. doi:10.1006/jasc.1994.1004

Porat, N., Chazan, M., Grün, R., Aubert, M., Eisenmann, V., & Horwitz, L. K. (2010). New radiometric ages for the Fauresmith industry from Kathu Pan, southern Africa: Implications for the Earlier to Middle Stone Age transition. *Journal of Archaeological Science, 37*(2), 269-283. doi:10.1016/j.jas.2009.09.038

Rakotoarivelo, A. R., O'Donoghue, P., Bruford, M. W., & Moodley, Y. (2019). An ancient hybridization event reconciles mito-nuclear discordance among spiral-horned antelopes. *Journal of Mammalogy, 100*(4), 1144-1155. doi:10.1093/jmammal/gyz089

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology, 67*(5), 901-904. doi:10.1093/sysbio/syy032

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology, 29*(1), 24-26. doi:10.1038/nbt.1754

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539-542. doi:10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution, 34*(12), 3299-3302. doi:10.1093/molbev/msx248

Simonsen, B. T., Siegismund, H. R., & Arctander, P. (1998). Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: High variation but low differentiation. *Molecular Ecology, 7*(2), 225-237. doi:10.1046/j.1365-294x.1998.00343.x

Smitz, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., . . . Michaux, J. (2013). Pan-African genetic structure in the African buffalo *Syncerus caffer*: Investigating intraspecific divergence. *PLoS One, 8*(2), e56235. doi:10.1371/journal.pone.0056235

Smitz, N., Cornélis, D., Chardonnet, P., Caron, A., Garine-Wichatitsky, M. d., Jori, F., . . . Michaux, J. (2014). Genetic structure of fragmented southern populations of African Cape buffalo (*Syncerus caffer caffer*). *BMC Evolutionary Biology, 14*, 203-222. doi:10.1186/s12862-014-0203-2

Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology, 56*(4), 564-577. doi:10.1080/10635150701472164

Tambling, C., Venter, J., du Toit, J., & Child, M. F. (2016). A conservation assessment of *Syncerus caffer caffer*. In M. F. Child, L. Roxburgh, E. Do Linh San, D. Raimondo, & H. T. Davies-Mostert (Eds.), *The Red List of Mammals of South Africa, Swaziland and Lesotho*. South Africa: South African National Biodiversity Institute and Endangered Wildlife Trust.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution, 3*(2), 170-177. doi:10.1038/s41559-018-0777-y

Taylor, W. A., Lindsey, P. A., & Davies-Mostert, H. (2016). An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa. *The Endangered Wildlife Trust*, Johannesburg.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics, 14*(2), 178-192. doi:10.1093/bib/bbs017

van Hooft, W. F., Groen, A. F., & Prins, H. H. T. (2000). Microsatellite analysis of genetic diversity in African buffalo (*Syncerus caffer*) populations throughout Africa. *Molecular Ecology, 9*(12), 2017-2025. doi:10.1046/j.1365-294X.2000.01101.x

van Hooft, W. F., Groen, A. F., & Prins, H. H. T. (2002). Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology, 11*(2), 267-279. doi:10.1046/j.1365-294X.2002.01429.x

van Hooft, W. F., Groen, A. F., & Prins, H. H. T. (2003). Genetic structure of African buffalo herds based on variation at the mitochondrial D-loop and autosomal microsatellite loci: Evidence for male-biased gene flow. *Conservation Genetics, 4*(4), 467-477. doi:10.1023/A:1024719231545

van Wyk, A. M., Dalton, D. L., Kotzé, A., Grobler, J. P., Mokgokong, P. S., Kropff, A. S., & Jansen van Vuuren, B. (2019). Assessing introgressive hybridization in roan antelope (*Hippotragus equinus*): Lessons from South Africa. *PLoS One, 14*(10), e0213961. doi:10.1371/journal.pone.0213961