

Colonialism in South Africa leaves a lasting legacy of reduced genetic diversity in Cape buffalo

Liam Quinn¹  | Genís Garcia-Erill¹  | Cindy Santander¹ | Anna Brüniche-Olsen¹ | Xiaodong Liu¹  | Mikkel-Holger S. Sinding¹  | Michael P. Heaton²  | Timothy P. L. Smith² | Patrícia Pečnerová¹ | Laura D. Bertola¹ | Kristian Hanghøj¹ | Malthe Sebro Rasmussen¹ | Deon de Jager³  | Hans R. Siegismund¹ | Anders Albrechtsen¹ | Rasmus Heller¹  | Ida Moltke¹ 

¹Department of Biology, University of Copenhagen, Copenhagen, Denmark

²USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, New Brunswick, USA

³Section of Molecular Ecology and Evolution, The Globe Institute, University of Copenhagen, Copenhagen, Denmark

Correspondence

Ida Moltke, Rasmus Heller and Anders Albrechtsen, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
Email: ida@binf.ku.dk; rheller@bio.ku.dk; albrecht@binf.ku.dk

Handling Editor: David Coltman

Abstract

The iconic Cape buffalo has experienced several documented population declines in recent history. These declines have been largely attributed to the late 19th century rinderpest pandemic. However, the effect of the rinderpest pandemic on their genetic diversity remains contentious, and other factors that have potentially affected this diversity include environmental changes during the Pleistocene, range expansions and recent human activity. Motivated by this, we present analyses of whole genome sequencing data from 59 individuals from across the Cape buffalo range to assess present-day levels of genome-wide genetic diversity and what factors have influenced these levels. We found that the Cape buffalo has high average heterozygosity overall (0.40%), with the two southernmost populations having significantly lower heterozygosity levels (0.33% and 0.29%) on par with that of the domesticated water buffalo (0.29%). Interestingly, we found that these lower levels are probably due to recent inbreeding (average fraction of runs of homozygosity 23.7% and 19.9%) rather than factors further back in time during the Pleistocene. Moreover, detailed investigations of recent demographic history show that events across the past three centuries were the main drivers of the exceptional loss of genetic diversity in the southernmost populations, coincident with the onset of colonialism in the southern extreme of the Cape buffalo range. Hence, our results add to the growing body of studies suggesting that multiple recent human-mediated impacts during the colonial period caused massive losses of large mammal abundance in southern Africa.

KEYWORDS

cape buffalo, conservation genetics, demographic history, inbreeding, population genetics

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Low genetic diversity is a cause for concern for the ongoing viability of many species, leading to an increased risk of compromised evolutionary potential (Frankham, 2005). However, for many species it is not known which factor(s) have caused their low genetic diversity. The drivers of genetic diversity in various wildlife species have been identified with the aid of genome-wide data, which has linked present-day diversity levels with a variety of biological traits and evolutionary events, including ecological versatility (Pečnerová et al., 2021), inbreeding (Prado-Martinez et al., 2013), disease epidemics (DeCandia et al., 2021), life-history traits and habitat availability (Brüniche-Olsen et al., 2021; Heller et al., 2010). Furthermore, it is well known that processes in a species' evolutionary history such as range expansions (Murray et al., 2017) and gene flow between populations (Fu et al., 2016), as well as recent anthropogenic pressures including habitat loss and hunting (Li et al., 2016; Morin et al., 2021) can impact species' genetic diversity. Moreover, studies have shown that the rate of decline in genetic diversity can be important in determining a species' future viability (Frankham, 2005; Kardos et al., 2021). Hence, understanding how demographic history has shaped levels of genetic diversity across wildlife populations is essential for informing conservation and population management of vulnerable species (Fuentes-Pardo & Ruzzante, 2017).

One near-threatened species whose levels of genetic diversity is likely to have been impacted by all of the above-mentioned factors is the African buffalo (*Syncerus caffer*), in particular its most widespread subspecies the Cape buffalo (*S. c. caffer*). The Cape buffalo is believed to have split from other African buffalo subspecies within the last 200,000 years (Smitz et al., 2013; Van Hooft et al., 2002). Since this split, two factors acting over extended evolutionary timescales are particularly likely to have played a role in present-day diversity levels. First, a gradual range expansion from its central African origin to southern Africa is believed to have taken place during the Late Pleistocene (100,000–20,000 years ago) caused by climatically induced expansion of grassland habitats (Van Hooft et al., 2002). Second, any subsequent changes in the environment could have led to a marked change in population sizes (de Jager et al., 2021; Heller et al., 2010; Smitz et al., 2013).

In more recent times, Cape buffalo populations have been greatly impacted by a host of human-mediated disturbances that have led to dramatic declines in population sizes, resulting in the need for population management strategies in parts of the Cape buffalo range (de Jager et al., 2021). These include human population expansions during the African Neolithic period (7000–1000 years before the present) (Heller et al., 2012) as well as European colonization since the 17th century with associated intensification of hunting, habitat fragmentation and culling of buffalo to protect livestock from diseases they are known to carry (Metzger et al., 2010). The introduction of novel pathogens as an indirect result of colonialism has had disastrous impacts on Cape buffalo populations. In the final decades of the 19th century, populations were reduced by as much as 90% across the species range due to the rinderpest pandemic (Metzger

et al., 2010; Njeumi et al., 2012; Young, 1969). Further population size decreases in recent decades have led to the African buffalo being categorized as 'Near-Threatened' on the IUCN Red List (IUCN, 2018). Despite this, previous studies based on markers such as microsatellites and mitochondrial DNA (mtDNA) have suggested surprisingly high levels of genetic diversity across the Cape buffalo range with estimates of lower values in some South African populations (de Jager et al., 2020; O'Ryan et al., 1998; Smitz et al., 2013, 2014; Van Hooft et al., 2000).

The first and so far only whole-genome-based study of Cape buffalo confirmed the presence of populations with high diversity, but also identified two populations with reduced levels (de Jager et al., 2021). However, that study only included samples from South Africa, and thus the genome-wide diversity of Cape buffalo outside of South Africa, as well as its drivers, remains unknown (O'Ryan et al., 1998; Smitz et al., 2013, 2014; Van Hooft et al., 2000).

Motivated by this, we analysed whole genome sequencing data from 59 Cape buffalo from eight populations across its range. Our primary goals were three-fold: first, to assess genetic diversity levels in Cape buffalo populations from across the subspecies' range based on genome-wide data; second, to investigate whether these genome-wide diversity levels were influenced by long-term phylogeographical and demographic changes, such as the presumed range expansion into southern Africa or other climatically driven changes in population size; and third, to evaluate the role of very recent demographic change in impacting genetic diversity, and thereby investigate the contribution of the rinderpest pandemic and other human-mediated disturbances that have impacted Cape buffalo habitats. By pursuing these goals we aimed to enhance our understanding of Cape buffalo genetic diversity and the factors that have affected it.

2 | MATERIALS AND METHODS

2.1 | Samples

Twenty-two samples of Cape buffalo skin tissue from Amboseli National Park (KY-Amboseli), Kenya, and previously analysed in Heller et al. (2010) were obtained from the University of Copenhagen. See Heller et al. (2010) for sampling details. Eight samples from OI Pejeta Conservancy Kenya (KY-Pejeta), Rungwe Tanzania (TZ-Rungwe) and Chewore Zimbabwe (ZM-Chewore) were provided by the USDA, ARS and U.S. Meat Animal Research Center (USMARC). Their hide samples were acquired from a licensed importer that legally imports hunter-killed, acidified, dried hides for commercial taxidermy. Other samples were sourced from existing publications (Table S1).

2.2 | DNA extraction

Following the manufacturer's protocol, the Qiagen DNeasy Blood & Tissue Kit (Qiagen) was used for DNA extraction. Subsequently,

RNase was added to the samples to ensure they consist of RNA-free genomic DNA. The DNA concentrations were then measured with a Qubit 2.0 Fluorometer and a Nanodrop before using gel electrophoresis to check the genomic DNA.

DNA from dried acidified Cape buffalo hide samples obtained by USMARC was extracted by first shaving the hair of a two-square-inch sample with a clean one-sided razor blade. Approximately 25 fragments of hide were shaved from the grain and chorion layer perpendicular to the hair side of the sample. Each shaved fragment was approximately the size of a human fingernail clipping. The shaved hide fragments were hydrated in phosphate-buffered saline overnight and DNA was extracted with a typical phenol/chloroform method and stored at 4°C in 10 mM TrisCl, 1 mM EDTA (pH 8.0) as previously described (Heaton et al., 2008).

2.3 | Sequencing

Sequencing libraries were built from the DNA extracts and sequenced by BGI China. The sequencing was performed using a 2 × 150-bp paired-end setup on the Illumina NovaSeq6000 S4 platform, targeting ~15× coverage. In total, the generated data comprised 729,130,136,100 bases.

For the USMARC samples, ~5 µg of Cape buffalo genomic DNA was used to make 500-bp paired-end libraries and sequenced as previously described (Kalbfleisch et al., 2018).

2.4 | Mapping

Overlapping read pairs were merged with `NGMERGE` before mapping with `BWA-MEM` (version 0.7.17; default settings) to a *Bos taurus* genome (BosTau9, GenBank accession GCA_002263795.2, representing a Hereford breed of cattle). Read duplicates (`markdup`, default settings) were marked and removed (`-F 3852`) using `SAMTOOLS` (version 1.9). We retained properly paired and mapped paired-end reads. All reads with mapping quality below 30 were excluded from all downstream analysis.

2.5 | Reference genome quality filtering

To obtain high-quality data we excluded data from parts of the genome using several different filters.

First, we used `REPEATMASKER` (version 4.08) to identify low-complexity and repeat regions in the genome using sensitive mode, and commands `-engine wublast -s -no_is -cutoff 255 -frag 2000` (Jiang et al., 2008). We also identified regions of the genome with an excess of heterozygosity. These can arise due to the presence of unknown paralogue genomic regions or repeats being represented only once in the reference genome, resulting in reads from multiple parts of the genome being mapped to a single location. To identify these

regions we used `PCANGSD` (Meisner & Albrechtsen, 2018) where we calculated per site inbreeding coefficients (F) in a way that accounts for population structure and missing genotypes. The first two principal components were used to account for population structure when obtaining individual allele frequencies. We identified regions as 10-kb windows containing sites that were found to have excessive heterozygosity ($F < -0.9$, $p < 10^{-6}$). Finally, we identified regions that showed excessively low or high sequencing depth across all samples. We first estimated the combined depth for all autosomal sites across all samples with `ANGSD` (Korneliussen et al., 2014), and then excluded sites with depth below 0.5× the median depth and sites with depth above 1.5× the median depth. Finally, we created a file of all sites in the genome that did not lie in any of the problematic regions identified above. This sites file was used in all subsequent analysis to filter out data from problematic regions of the genome.

2.6 | Sample quality filtering

We also performed several different filtering steps at sample level described below.

2.6.1 | Error rates estimation

To identify and exclude highly error-prone samples from the data set, we estimated error rates in `ANGSD` using the “perfect-individual” approach as described in Orlando et al. (2013). Sample AB-SA was selected as the “perfect-individual” sample for the analysis based on the overall mapping statistics and sequence depth. A consensus sequence was generated for this sample with the `-doFasta 2` option in `ANGSD`, taking the most commonly observed base (`-doAncError 2`) as the consensus. Since the “perfect-individual” approach assumes equal genetic distances of the samples to the outgroup, we used the `bostau9` reference genome as the ancestral (outgroup) sequence. Error rates were then measured as an excess/deficit of derived alleles from the outgroup compared to this “perfect individual” sample. Both for generating the consensus sequence and for the error estimation, we restricted the data to a base quality of at least 30. No samples were removed after assessing error rates in the data set.

2.6.2 | Identification of related and duplicated samples

We removed closely related samples from analyses. These were identified with `PLINK 1.9` using the relatedness coefficient (`2×kinship coefficient`) (`--genome`). For each pair with a relatedness coefficient estimate above 0.2 we removed the sample with the highest error rates inferred by `ANGSD`. This analysis was performed separately for the three populations where sufficient sample sizes

were available for population allele frequency estimation (KY-Amboseli, RSA-HiP, RSA-KNP). Seven Amboseli samples were removed, leaving no pairwise relationships with relatedness coefficient above the threshold. For the KY-Pejeta, RSA-AENP, RSA-MNP and ZM-Chewore populations, for which the low sample size prevents reliable estimation of allele frequencies, we estimated relatedness based on the allele frequency-free method `IBSRELATE` (Waples et al., 2019). We used `ANGSD` and `REALSFS` to estimate the two-dimensional site frequency spectrum [2D-SFS] between each pair of samples from each population, from which we calculated the R_0 , R_1 and `KING` statistics. Based on them, we found a duplicate pair of samples from ZM-Chewore and we excluded one of the samples.

2.7 | Genotype imputation, genotype calling and genotype likelihoods

From the mapped read data we first removed reads with low mapping quality and bases with a low-quality score: `-minQ 30 -minMapQ 30`. For some analyses we needed information for all sites (e.g., pairwise sequentially Markovian coalescent [`psmc`]) and for other analysis we also needed data for common variable sites.

To obtain genotype calls for common variable sites we performed haplotype imputation. We first called SNPs in Cape buffalo using `ANGSD` from genotype likelihoods, with a minimum p -value of 10^{-6} and a minor allele frequency [MAF] above 0.05. We used the `GATK` genotype likelihood model (McKenna et al., 2010). We then used the genotype likelihoods for the variable sites as input for `BEAGLE` version 3 (Browning & Browning, 2011) to perform haplotype imputation jointly on all samples. We excluded sites with an imputation score (estimated R^2) below 0.9 and called the genotype with the maximum posterior probability in all remaining sites, removing sites where any sample had a maximum posterior probability lower than 0.95. We will refer to this data set as the imputed data, which was used for most analyses (Table S2).

For a few of the analyses that are very sensitive to genotyping errors using the imputed data was not appropriate. Specifically, for the `psmc` analysis we called genotypes separately for nine high-depth samples using `BCFTOOLS` (version 1.9) and `-c` for calling genotypes. We used a minimum depth per site selecting the minimum of either six reads or a third of the genome average depth. A maximum depth per site equal to twice the genome-wide average depth was used. A minimum of two reads supporting each allele were required to call a heterozygous genotype. Similarly, for identification of long runs of homozygosity (ROH) in each sample we made high-quality genotype calls by jointly calling them using `BCFTOOLS` and then keeping only genotypes supported by a minimum read depth of 10 and requiring a minimum of three alternative and reference reads to call a heterozygous genotype. Unlike `psmc`, long ROH can be identified despite the ascertainment bias. Finally, for the estimated heterozygosity, which is extremely affected by ascertainment bias, we used the genotype likelihoods data instead

of called genotypes. For a full overview of the use of the different data sets see Table S2.

2.8 | Principal components analyses (PCAs)

We performed PCA on the 59 not closely related Cape buffalo based on the imputed data using `PLINK` 1.9 with the `--pca` command. The data were pruned for sites in linkage disequilibrium using the `"-indep-pairwise"` command with a window size of 50 variants, a step size of five variants and pairwise r^2 threshold of 0.2.

2.9 | Linkage disequilibrium (LD) curves

LD curves were generated for six Cape buffalo populations with four individuals from each population to allow equal sampling. The curves were computed using the `"LDdecay"` function in the `"LDdecay.R"` package in R (Albrechtsen et al., 2009; Korneliusson et al., 2014) using the imputed data in `plink` format.

2.10 | F_{ST} estimation

We estimated F_{ST} between Cape buffalo populations by applying the Hudson's estimator (Bhatia et al., 2013) to the imputed data. Individuals were grouped in populations according to their sampling location.

2.11 | ADMIXTURE analyses

We estimated admixture proportions for 59 individuals using `ADMIXTURE` (Alexander et al., 2009) on the imputed data to identify any admixed individuals within sampling locations due to recent translocations. `ADMIXTURE` was run from $K = 2$ to $K = 6$. For each K we performed several independent optimization runs until convergence. Convergence was not achievable for all K . For the runs that converged we assessed the model fit by estimating the pairwise correlation of residuals with `EVALADMIX` (García-Erill & Albrechtsen, 2020). One individual in the KY-Amboseli population was shown to have an `ADMIXTURE` profile most closely resembling two individuals from KY-Pejeta. This was visible from $K = 4$ to $K = 6$; this sample was removed from all subsequent analyses.

2.12 | Estimation of genome-wide heterozygosity

We estimated heterozygosity for each sample using `ANGSD` after obtaining a per-individual site frequency spectrum (SFS) using the `REALSFS` program within `ANGSD` (`-dosaf 1 -gl2 -minQ 30 -minMapQ 30`). Using genotype likelihoods takes into account uncertainty in the data, which was necessary given that many of the samples in

this data set are of low average sequencing depth. In addition to our 58 Cape buffalo samples we also estimated heterozygosity for nine additional samples. These included three swamp buffalo, five water buffalo and a cattle sample.

2.13 | TREEMIX

From the imputed data we estimated allele counts for each population, which we used as input for TREEMIX (Pickrell & Pritchard, 2012). We ran TREEMIX assuming no migrations, grouping SNPs in blocks of 687 SNPs, and setting a pair of individuals from the Italian water buffalo population as an outgroup.

2.14 | Identity by state (IBS) neighbour-joining tree

We created a neighbour-joining tree using the imputed data set in PLINK. Using the commands “-distance square ibs” and “allele-ct” pairwise IBS values were calculated for all individuals which included 58 Cape buffalo samples and two water buffalo samples intended to act as an outgroup. In R using the matrix of pairwise IBS distances as input, the APE plotting package created an unrooted neighbour-joining with the “ape::(nj)” command.

2.15 | Admixture graphs

To estimate admixture graphs for the eight Cape buffalo populations, we used the “find_graphs” function in ADMIXTOOLS2 (Maier et al., 2022) on the imputed data set using the Italian water buffalo population as an outgroup. For each allowed number of admixture events, we ran “find_graphs” 300 times to obtain test scores for each admixture event and kept the admixture graph with the highest test score. The test score was obtained by optimizing a topology from a subset of the *f*-statistics and testing on the remainder. We then selected the highest scoring admixture graph for each admixture event.

2.16 | Estimation of population split times

We applied the “Two-Two” (TT) method (Sjödén et al., 2021) to estimate split times between populations. This method estimates pairwise population split times using one sample from each population, by fitting a coalescence model to the 2D-SFS between the two samples. The model assumes a constant ancestral population size and no gene flow after the split, but it avoids any assumptions on the population sizes after the split. We obtained the 2D-SFS for each population pair using called genotypes from each of the highest depth samples from each population used (average sequencing depth > 10×). We polarized the ancestral state as the allele in the bosTau9 reference, requiring that the reference allele was also

observed in homozygous state in the impala and lesser kudu samples, and excluded from the analyses those sites where this was not the case.

2.17 | Gene flow analysis using *f*-branch statistics

Gene flow between eight different Cape buffalo populations was investigated using the *f*-branch metric, a statistic related to the *f*-4 statistic, which we calculated by applying the DSUITE software package (version 0.3; Malinsky et al., 2021) to the imputed data set. To do so, we first used Dtrios, a part of DSUITE, to calculate *D*-statistics for all possible trio combinations given the topology of the highest scoring admixture graph with two admixture events depicted in Figure 3. Then we applied Fbranch, another part of DSUITE, to these *D*-statistics in order to calculate the *f*-branch statistics.

2.18 | PSMC

We used PSMC (Pickrell & Pritchard, 2012; Li & Durbin, 2011) to estimate the population size trajectory back through time for our 10 high-coverage Cape buffaloes. PSMC was applied with default settings. For scaling, we used a mutation rate of 1.5×10^{-8} and generation time of 7.5 years based on a value obtained from previous studies of Cape buffalo (de Jager et al., 2021). All samples were analysed using their original depths (Figure S7B), and then samples were downsampled to 10.3× average genome sequencing depth to ensure equal coverage using SAMTOOLS with the -b and -s filters with -s set to the proportion of reads required to equal 10.3× average genome sequencing depth per sample.

2.19 | POP SIZE ABC

We used POPSIZEABC (Boitard et al., 2016) on the imputed data to estimate changes in population sizes in all samples from three of the Cape buffalo populations where sufficient numbers of samples were available ($n > 12$). POPSIZEABC takes vcf files per chromosome as an input. From this it estimated LD curves and SFS for tested populations. A vcf file was created from the imputed data set and the three populations with sufficient sample sizes were analysed. The same recombination rate and mutation rate used in PSMC were used in POPSIZEABC. POPSIZEABC also requires multiple simulations of demographic scenarios to compare a posterior distribution of simulation derived parameters to those observed in the real data. For each population, 210,000 simulations were performed for 100 2-Mb regions per simulation as per the suggested settings in the publication for the software. A minimum MAF threshold of 0.1 was applied to calculation of the SFS and 0.2 for calculation of the LD curves, again in accordance with the suggested parameters in the POPSIZEABC publication.

2.20 | Runs of homozygosity analysis

Runs of homozygosity (ROH) were inferred based on strict genotype calling filters (see genotype imputation, genotype calling and genotype likelihoods) where heterozygous genotypes have at least three reads supporting each allele. This minimizes the likelihood of sequencing errors fragmenting long ROH.

We used the command “-homozyg” in PLINK version 1.9 with default settings, except allowing a maximum of three heterozygous calls and 20 missing calls when detecting ROH (--homozyg-window-het 3 --homozyg-window-missing 20). Minimum ROH length was set at the default of 1 Mb.

Fractions of the genome as ROH were calculated by dividing the total length of ROH per sample by the entire genome length. Using a published approach (Meyermans et al., 2020) to estimate the entire genome length per sample, “pseudo-homozygous” individuals were created with each individual's genotype fixed to homozygous for all sites. The same PLINK ROH command was rerun and the total length of ROH from these individuals was taken as the total genome length.

Genome-wide heterozygosity was then recalculated for each sample taking into account these extended portions of the genome found to be homozygous. This ROH-adjusted heterozygosity was calculated by: ROH-adjusted heterozygosity = Heterozygosity/(1-fraction of genome which has ROH).

We converted the ROH tract lengths to generation estimates using an approach from a previous publication (Saremi et al., 2019). We used an estimated average recombination rate from cattle of 1.275 centimorgans/Mb (Saremi et al., 2019; Weng et al., 2014) and the equation $g = 100/(2rL)$, where g is the time in generations, r is the recombination rate and L is the length of the ROH tract in megabases.

3 | RESULTS

We generated whole genome sequencing data from 22 Cape buffalo individuals from Amboseli National Park in Kenya (KY-Amboseli), two from Ol Pejeta Conservancy Kenya (KY-Pejeta), five from Chewore Lodge Zimbabwe (ZM-Chewore) and one from Rungwe East Game Reserve in Tanzania (TZ-Rungwe), with a depth of coverage 8–21× per individual. In addition, we obtained sequencing data for 41 Cape buffalo samples from publicly available data (depth of coverage 6–32×) from four other localities in South Africa: Kruger National Park (RSA-KNP), Mokala National Park (RSA-MNP), Hluhluwe-iMfolozi Park (RSA-HiP) and Addo Elephant National Park (RSA-AENP) (see Figure 1a for map of samples and Table S1 for sample information). Finally, for a broader comparison we included eight publicly available samples from three water buffalo (*Bubalus bubalis*) populations with both the river and swamp subspecies included. After mapping to the cattle reference genome and sample quality filtering, 59 Cape buffalo remained with 12 samples discarded due to a close familial relationship or failing quality control procedures (see Table S2 for a summary of data used in each analysis and methods for relatedness and quality control of samples).

3.1 | Evaluation of population units

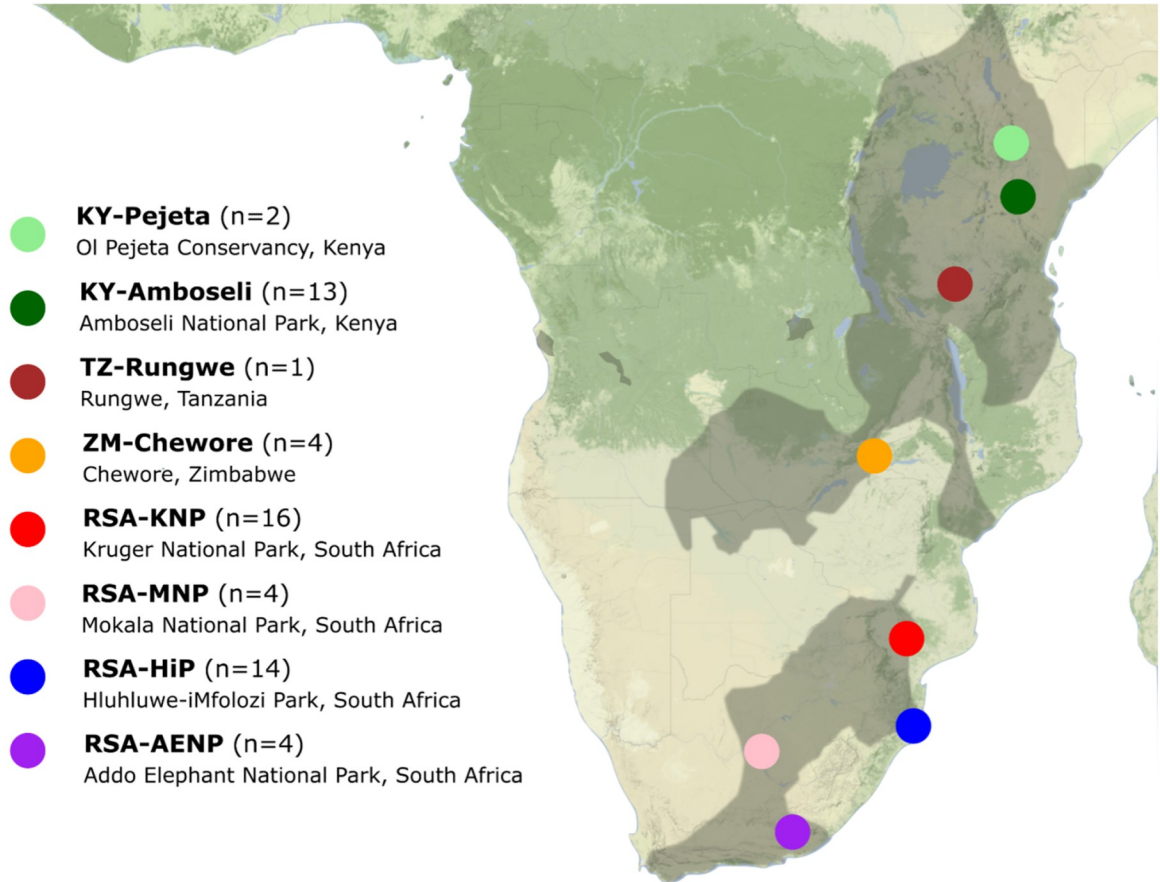
First, to ensure the sampling localities constitute evolutionarily meaningful population units for further analyses, we performed a series of population structure analyses. A PCA shows samples from the same sampling location clustering together in the first two principal components, forming nonoverlapping clusters except for two South African populations, RSA-KNP and RSA-MNP (Figure 1b). The latter is unsurprising, given that the RSA-MNP population was established in 1999 from individuals translocated from RSA-KNP in order to create a disease-free population (de Jager et al., 2021). In line with this, the estimated F_{ST} between them is also the lowest observed (0.01) (Table S3). The remaining F_{ST} values range from relatively low between populations across the Cape buffalo distribution from Kenya to RSA-KNP (average 0.054), to much higher between the two southernmost localities (RSA-HiP and RSA-AENP) and other localities (average 0.21). An ADMIXTURE analysis (Figure 1d; Figure S1) supported the clustering of the individuals into groups according to sampling location, with the exception of the two locations with low sample sizes (KY-Pejeta and TZ-Rungwe). All samples in these two locations appear admixed, but this is likely an artefact of low sample size (Lawson et al. 2018) rather than actual admixture, since EVALAD-MIX shows a poor model fit for these two populations even at $K = 6$ with elevated correlation of residuals (Figure S1). More surprisingly, the RSA-KNP samples are assigned ancestry from two clusters and not just one. However, F_{ST} between these two clusters is very low (0.02) suggesting they are genetically very similar.

In the PCA, one of the samples from KY-Amboseli is placed between the other KY-Amboseli samples and the KY-Pejeta samples, suggesting it may be admixed. This was further supported by the admixture analysis and this sample was therefore excluded from all subsequent analyses to avoid recent admixture affecting downstream inferences (Gopalan et al., 2022). To perform an additional check that there were no additional signs of recent admixture in any of the populations, we generated an LD decay curve for each of the populations (Figure 1c; Figure S2). The LD curves converge to the same level as expected in panmictic populations, ruling out very recent admixture. Based on these results, we decided to continue our analyses with all the remaining samples and treat each sampling locality as a population unit. RSA-KNP and RSA-MNP could in principle be pooled into a single population given how genetically similar they are. However, we chose to keep them as distinct populations as we know they have had separate histories for the past few generations.

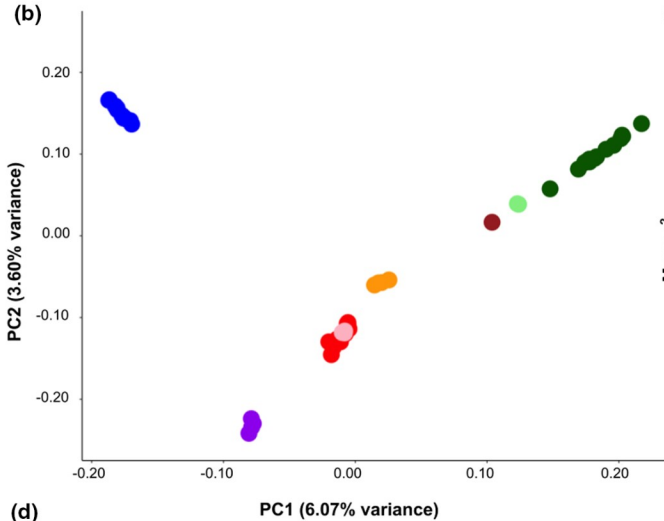
3.2 | Genome-wide heterozygosity levels

We estimated the genome-wide heterozygosity for each sample in order to compare differences in genetic diversity between the eight Cape buffalo populations and three water buffalo populations (Figure 2). Overall, the Cape buffalo had high genome-wide heterozygosity levels across its range compared to the three water buffalo populations as well as other large mammals estimated using

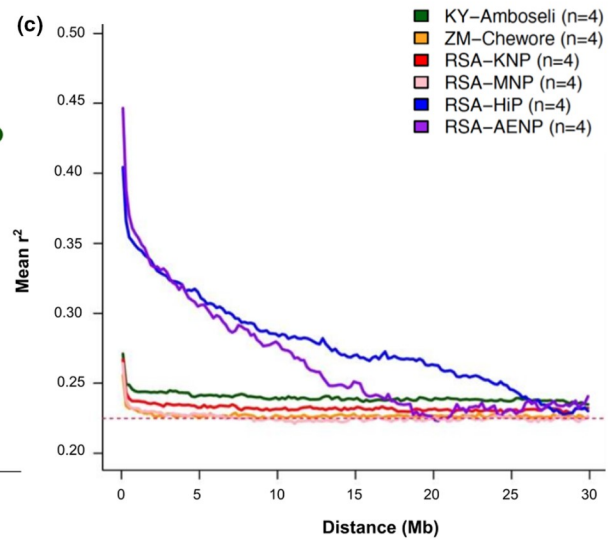
(a)



(b)



(c)



(d)

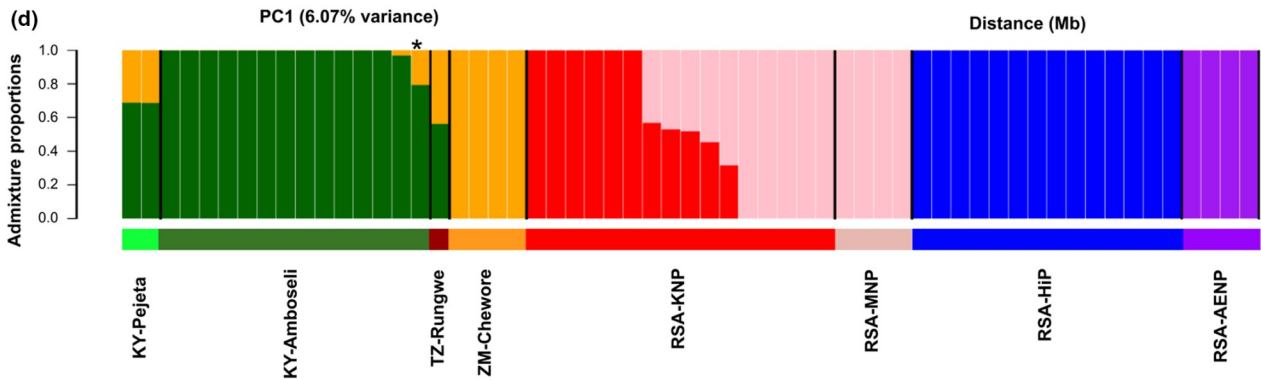


FIGURE 1 (a) Overview of Cape buffalo sampling locations with numbers of samples for each population given in parentheses and the Cape buffalo distribution shown by grey shading (source <http://www.iucn.org>). (b) PC1 and PC2 for all 59 Cape buffalo samples coloured by their sampling location. (c) Linkage disequilibrium (LD, measured as mean r^2) for pairs of SNPs stratified by their genomic distance for the eight Cape buffalo populations estimated from four samples from each population. To better illustrate convergence a dashed line has been added around the lowest observed mean r^2 (d) Admixture proportions estimated for 59 Cape buffalo using ADMIXTURE assuming the presence of six ancestral populations ($K = 6$). We removed the individual marked with an asterisk from subsequent analysis as it appears admixed.

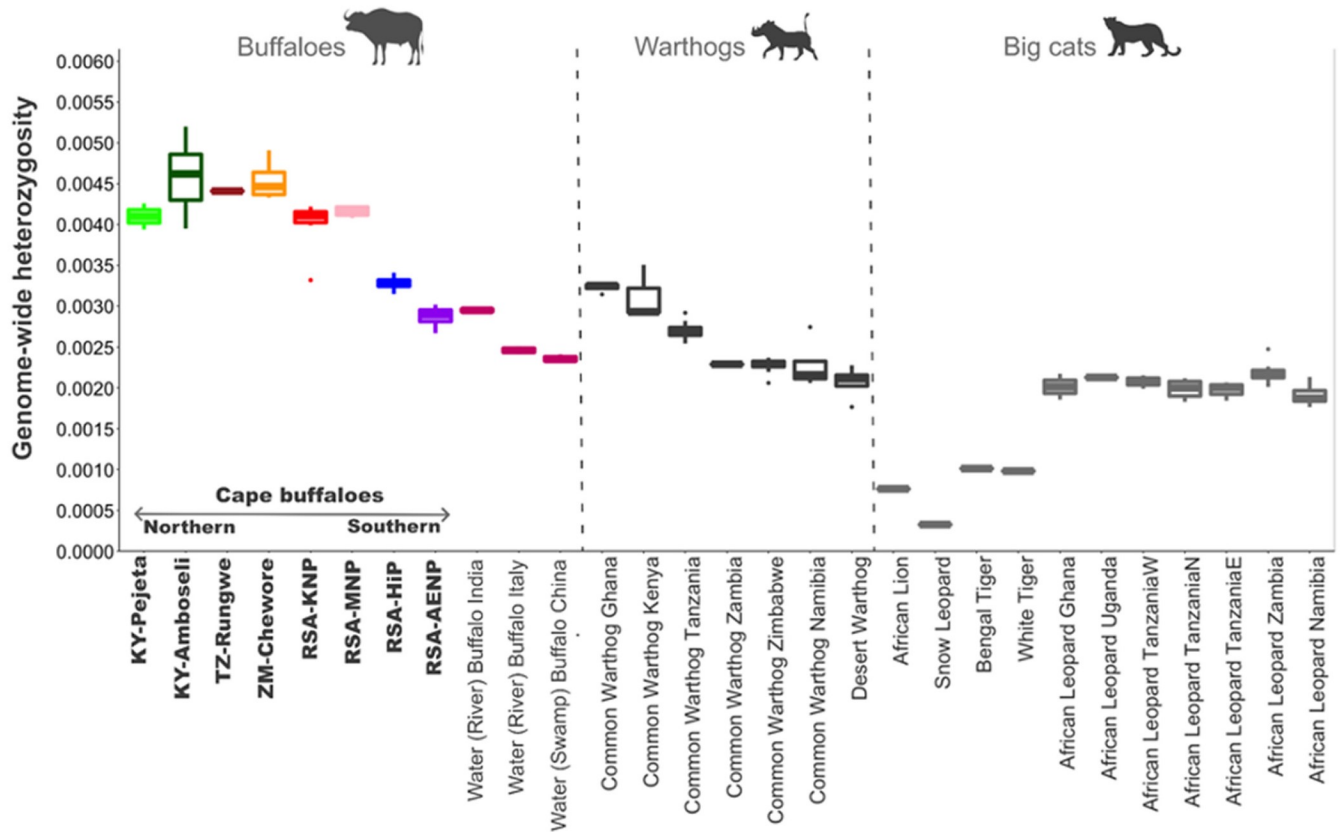


FIGURE 2 Genome-wide heterozygosity levels estimated for 58 Cape buffalo samples as well as samples from three water buffalo populations (two river buffalo populations and one swamp buffalo population) accompanied by estimates from mammal species that have been previously analysed using the same pipeline (García-Erill et al., 2022; Pečnerová et al., 2021), and thus are directly comparable. The estimates are shown stratified by population with a boxplot for each, where the central horizontal line shows the median, the box shows the 25th and 75th percentiles, and the whiskers extend to the largest/smallest value, but only until 1.5x the interquartile range, beyond which points are plotted as outliers. Warthog icon credit: PhyloPic/Jan A. Venter, Herbert H. T. Prins, David A. Balfour & Rob Slotow (vectorized by T. Michael Keesey), shared under Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>) without changes.

the same method and site filtering. Mean genome-wide heterozygosity was highest in the KY-Amboseli population, but remained above 0.4% in all except the two southernmost populations, RSA-HiP and RSA-AENP. In fact, mean genome-wide heterozygosity of RSA-AENP is comparable to that of the domesticated water buffalo.

We also estimated genome-wide diversity (Tajima's π) for each of the populations with a sufficient number of samples ($n > 12$). The estimated π values were very similar to the mean heterozygosity for the same populations, suggesting that the observed differences are not an artefact of the genetic diversity measure used (see Table S4).

3.3 | Overall phylogeography in Cape buffalo

To investigate which factors have affected the observed heterozygosity patterns in Cape buffalo, we first considered the possible contribution of the presumed gradual range expansion of Cape buffalo from their Central African origin to southern Africa during the Late Pleistocene with populations splitting off one at a time along the range (Lorenzen et al., 2012). Interestingly, we did not observe a linear decrease in heterozygosity southward along the range, which is the expected signature of such a range expansion (Goodsman et al., 2014). For instance, heterozygosity is higher in ZM-Chewore (Zimbabwe)

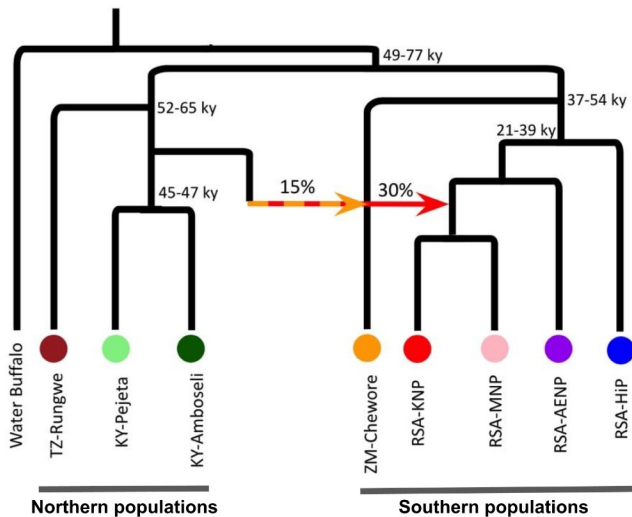


FIGURE 3 Admixture graph of population split and gene flow events in Cape buffalo. The tree topology and the gene flow events shown are based on the highest scoring admixture graph assuming two admixture events. The branch lengths do not reflect time or drift. Numbers at nodes are the ranges of split times in thousands of years between populations based on their positions in the tree estimated using the TT dating method.

than in KY-Pejeta (Kenya). To look further into this we inferred a population tree using TREEMIX (Figure S3). Interestingly, we obtained trees that did not support a simple gradual southwards range expansion. Instead, the TREEMIX tree shows an early split into a northern and a southern clade. To corroborate this finding, we used two other tree-based methods to infer the population tree: an IBS-based neighbour-joining tree approach and admixture graphs which allow for gene flow. Both the IBS-based tree (Figure S4) and the best scoring admixture graphs (Figure 3; Figure S5) showed the same initial split into a northern and a southern clade (for split times between these and other clades estimated using the coalescent-based TT method; see Figure 3 and Table S5). Moreover, the best scoring admixture graphs, when allowing for gene flow events, suggested several potential admixture events in the history of the populations, including a significant north to south gene flow into ancestral RSA-KNP, RSA-MNP and ZM-Chewore populations from a northern source. This suggests that there has been some connectivity between populations after a north-south split which is also indicated by *f*-branch statistics (Figure S6).

Collectively, these results suggest a more complex phylogeographical history of the Cape buffalo than the simple gradual range expansion that was previously proposed.

3.4 | Ancient and recent demographic history

Next, we investigated the connection between ancient population size changes (e.g., due to environmental changes) and present-day genetic diversity. For this we analysed six high-depth samples, representing six of the eight study populations, using the *p*SMC to infer

population size dynamics during the last 1 million years. We did not include any samples from RSA-AENP or RSA-MNP due to none of the samples from these populations having sufficient sequencing depth for this type of analysis. Following preliminary analyses suggesting a sensitivity of the *p*SMC results to sequencing depth (Figure S7A), the six samples were downsampled to have the same average sequencing depth to ensure comparability. The *p*SMC results show similar population size trajectories across locations (Figure 4a), with no indication of differences in genetic diversity being driven by significant differences in population size trajectories in the period between 30,000 and 100,000 years ago. The most ancient parts of the demographic histories for the samples from TZ-Rungwe and ZM-Chewore appear distinct from the other four samples, but this difference disappeared when the full depth data were analysed using *p*SMC, suggesting it is an artefact of low sequencing depth (Figure S7B).

To investigate whether more recent events have affected the observed genome-wide heterozygosity levels, we carried out another population size estimation analysis using *POPSIZEABC*, which has better resolution for recent population size changes. Specifically, we analysed the three study populations with sufficiently large sample sizes ($n > 12$), of which two had high genome-wide heterozygosity (KY-Amboseli and RSA-KNP) and one had low genome-wide heterozygosity (RSA-HiP). All three populations show a steady decline in population size beginning 8000–10,000 years ago and continuing until the present (Figure 4b). RSA-HiP shows a continued, drastic decline in population size in the last 500 years. Interestingly, such a drop is not present in KY-Amboseli or RSA-KNP, despite both of these populations being heavily impacted by the rinderpest pandemic at the end of the 19th century. The dramatic, recent population decline is consistent with the extended LD that was observed in the RSA-HiP and absent in KY-Amboseli and RSA-KNP (Figure 1c).

To further investigate the potential effect of recent population size declines we identified ROH longer than 1 Mb across all samples. ROH longer than 1 Mb are suggestive of inbreeding that took place within the last three centuries (see Table S6 for estimations of time origins of inbreeding tracts) and, interestingly, the average fraction of the genome that contains such ROH is low in most Cape buffalo populations (<5%) with the exception of RSA-HiP (19.9%) and RSA-AENP (23.7%) (Table S1). Moreover, of the total length of ROH that we identified in RSA-HiP and RSA-AENP, 58.3% and 56.7%, respectively, are from ROH longer than 5 Mb (Figure 5a). In fact, all individuals in RSA-HiP and RSA-AENP have inbreeding tracts longer than 10 Mb. This suggests that significant inbreeding took place within the last four to eight generations, and thus within the last century.

Importantly, we observed a clear correlation between genome-wide heterozygosity and the estimated total length of ROH in each sample (Figure S8A). Moreover, when we adjusted genome-wide heterozygosity by removing sites contained in the ROH regions longer than 1 Mb, we observe similar levels of heterozygosity in the RSA-HiP and RSA-AENP populations as in the other populations (Figure 5b). This suggests that recent inbreeding that occurred within the last three centuries is the main driver of the observed variability in genome-wide heterozygosity.

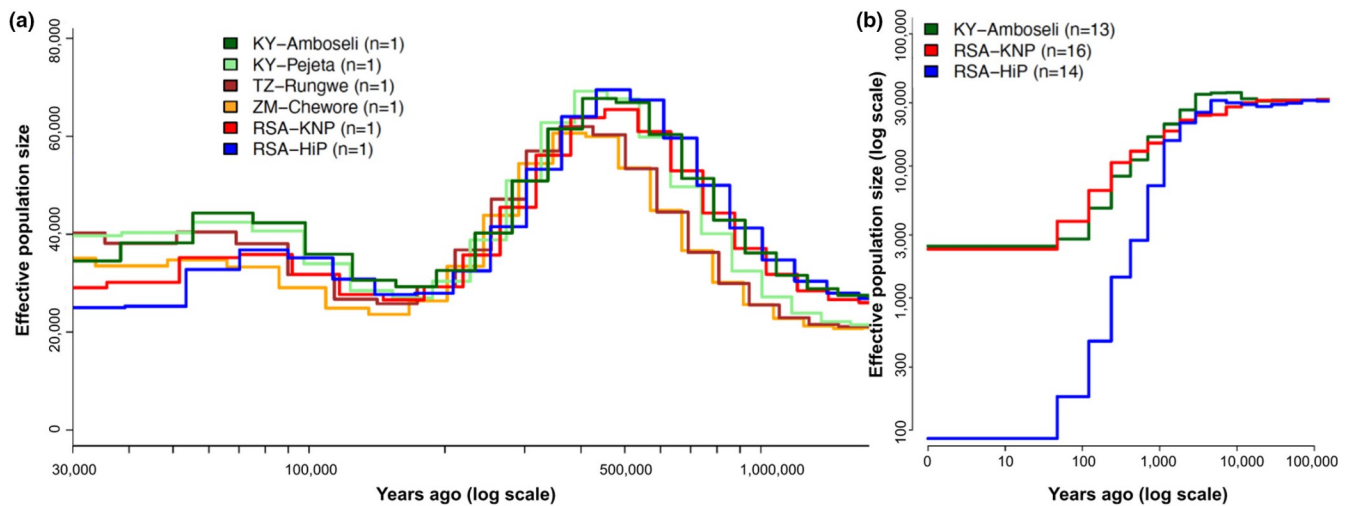


FIGURE 4 (a) Population size over time in the ancient past estimated by applying PSMC (mutation rate of 1.5×10^{-8} and generation time of 7.5 years) to six Cape buffalo samples downsampled to equal depth of 10.3 \times . Results without downsampling are shown in Figure S7B. Note that we chose not to display PSMC -estimated generated trajectories to more recent than 30,000 years since PSMC is not suited to estimate recent population sizes that are very noisy between samples. (b) Population size over time in the more recent past estimated by applying POPSIZEABC to data from three populations with a sufficient number of samples ($n > 12$).

4 | DISCUSSION

In this study we provide the first range-wide whole-genome analysis of the Cape buffalo. We conducted several previously unperformed analyses in this subspecies relating to its population structure, phylogeography and demography with a focus on understanding the landscape of genetic diversity.

We find that genetic diversity is generally high throughout the Cape buffalo range (Figure 2). This corroborates results based on other genetic markers (O'Ryan et al., 1998; Smits et al., 2013) as well as a recent study based on whole genomes from South African populations (de Jager et al., 2021). Furthermore, by adding samples outside South Africa we find that genetic diversity in two South African populations, RSA-HiP and RSA-AENP, is not only exceptionally reduced compared to other South African populations (de Jager et al., 2021; O'Ryan et al., 1998), but also compared to other populations throughout the Cape buffalo range. And importantly, unlike the previous studies we explicitly link this reduced genetic diversity to local demographic histories at different timescales (Figures 4 and 5). Specifically, our results are inconsistent with locally reduced diversity in these two populations being primarily caused by older demographic events, such as founder events in connection with the Cape buffalo range expansion. This is in contrast to what was found for another roughly codistributed African mammal, the common warthog (Garcia-Erill et al., 2022). Moreover, no substantial differences in ancient demographic history were observable between the RSA-HiP population and other populations with higher genome-wide heterozygosity. Thus, demographic differences over evolutionary timescales, caused for example by environmental changes, are an unlikely explanatory factor for the reduced heterozygosity. Our analyses also suggest that Cape buffalo phylogeography may be more complex than previously assumed (Lorenzen et al., 2012). For instance, several

of our analyses do not suggest a simple gradual southward range expansion, but rather a primary split with the formation of northern and southern populations followed by subsequent gene flow (Figure 3). Notably, our range of estimates for the major northern and southern (49,000–77,000 years ago, Figure 3) population split times are in agreement with previous single locus-based estimates of the range expansion of Cape buffalo (80,000 years ago: Heller et al., 2012; and 48,000 years ago: Smits et al., 2013), but we caution that our estimates could be affected by gene flow (Sjödén et al., 2021). We also note that either connectivity was high throughout most of the subspecies range, or genetic drift was low, as indicated by relatively low genetic differentiation ($F_{ST} = 0.09$) between Kenyan and the RSA-KNP South African population.

Whereas ancient demographic and phylogeographical processes did not seem to elicit marked differences in local genetic diversity (Figure 4a), we found pervasive support that recent declines in population size starting around 500 years ago and accelerating within the last century did to a large extent shape such differences. Specifically, we identified a significant recent population size decline in the RSA-HiP population relative to two other buffalo populations in the interval 100–1000 years ago (Figure 4b). This was strongly supported by elevated overall inbreeding levels in both RSA-HiP and RSA-AENP populations (Figure 5). In particular, a majority of ROH in these two populations were in ROH regions longer than 5 Mb, and all RSA-HiP and RSA-AENP individuals had ROH longer than 10 Mb. This suggests that inbreeding took place as recent as within the last century. We also show that recent inbreeding was in fact the main determinant of genome-wide heterozygosity differences across the Cape buffalo range (Figure 5a), as the majority of heterozygosity differences were attenuated after excluding the regions with ROH longer than 1 Mb. We do note that there are some remaining heterozygosity differences between populations after this correction

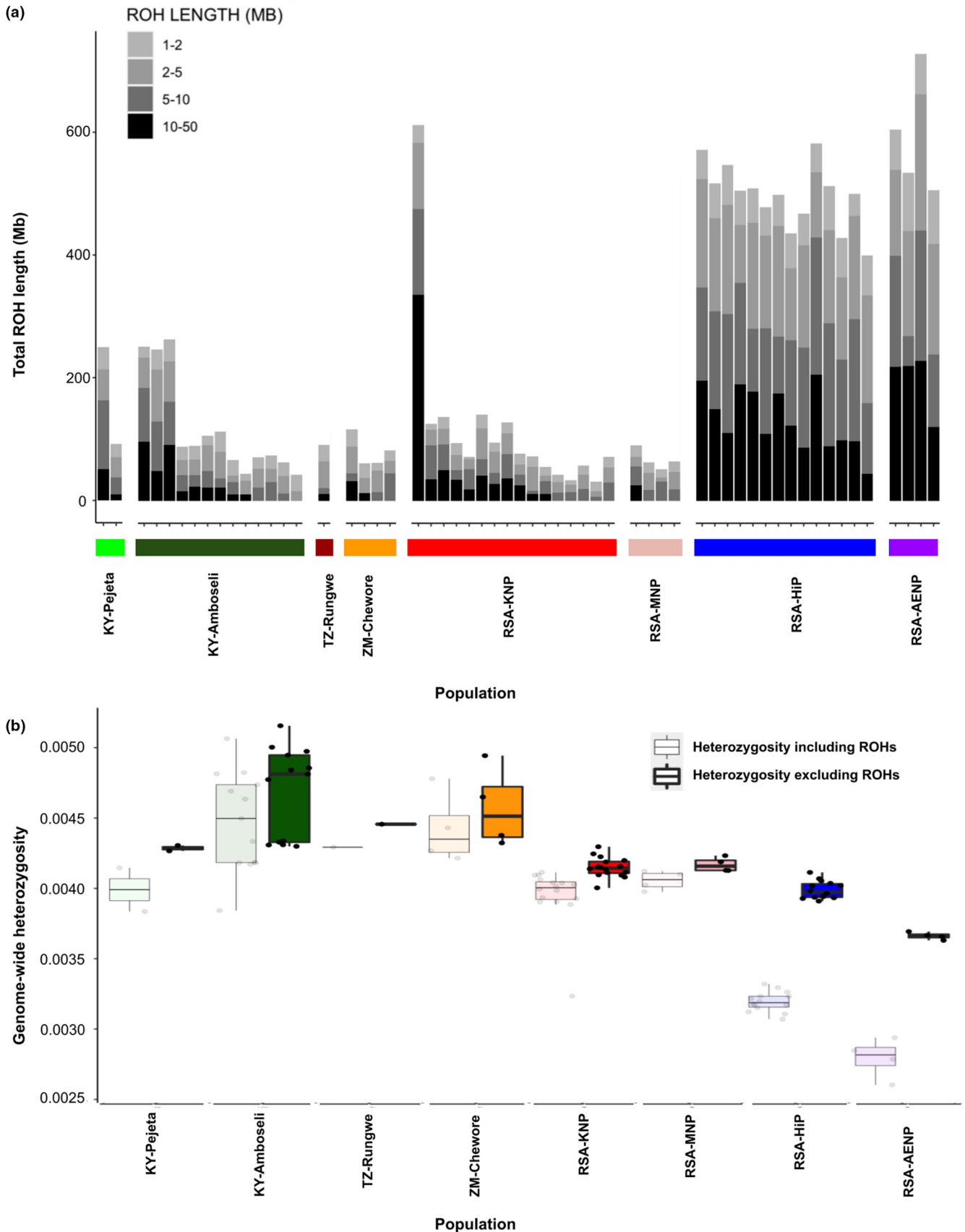


FIGURE 5 (a) Total length of ROH per individual categorized by ROH length. Longer ROH are indicative of more recent inbreeding events. (b) Genome-wide heterozygosity estimates adjusted for recent inbreeding via the fraction of ROH for each Cape buffalo genome (transparent boxes and points) with original heterozygosity estimates from Figure 2 placed alongside (opaque boxes and points) for comparison.

for recent inbreeding, possibly explained by other demographic processes including the southwards expansion. However, these differences are minor and require more data to verify. We therefore conclude that recent and progressive reductions in population size are the main explanation for why the southernmost populations have unusually low genome-wide diversity compared with other Cape buffalo populations.

There are two plausible external causes for recent population size changes in the Cape buffalo: the late 19th century rinderpest pandemic and direct human impacts, including habitat loss, hunting and culling. To our knowledge there is no indication that Cape buffalo in the southernmost part of the range suffered higher mortalities due to rinderpest than elsewhere on the continent. Therefore, we would expect similar levels of heterozygosity across the range if the loss of heterozygosity was mainly due to the rinderpest pandemic. Also, our analysis of recent demographic history suggests that the accelerated reduction in effective population size in the southernmost populations started considerably earlier than the rinderpest epidemic and continued afterwards, although the temporal resolution of any method inferring demographic history from genetic data is limited (Schraiber & Akey, 2015). Based on these two observations it does not seem plausible that rinderpest is the main cause. With regard to the other plausible cause, then interestingly there is a rich historical literature asserting that many large ungulate species declined rapidly following the establishment of the European Cape colony in present-day South Africa in 1652 (Van Sittert, 2005). These declines were caused by a combination of excessive hunting, fragmentation and reduction of habitat, as well as systematic eradication of wildlife, which was in some cases considered a nuisance for livestock keeping (Carruthers, 2005). Novel pathogens being introduced via livestock may also have contributed to these wildlife declines (Ndow et al., 2019), as in the case of rinderpest which entered the continent in the Horn of Africa (Tizard, 2021). As a direct consequence of this concentration of factors in southern Africa, several large mammals were severely reduced in numbers, examples include the extinction of the bluebuck (*Hippotragus leucophaeus*) (Hempel et al., 2021) and the quagga (*Equus quagga quagga*) (Leonard et al., 2005) and the near-extinction of the bontebok (*Damaliscus pygargus pygargus*), black wildebeest (*Connochaetes gnou*) (D'Amato et al., 2013) and black rhinoceros (*Diceros bicornis*) (Moodley et al., 2017). These human-mediated impacts are all documented to have played a role in decimating Cape buffalo populations in recent centuries (Van Hooft et al., 2000). In contrast, European colonization did not produce a similarly high impact on local wildlife until much later elsewhere in Africa with the establishment of the East African Protectorate in modern Kenya in 1895 (Rashid, 2014), the formation of German East Africa which covered modern-day Tanzania in 1890 (Pierskalla et al., 2019) and the arrival of the British South Africa Company in Zimbabwe in 1890 (Summers, 1994). Overall, it appears likely that the cause of the exceptionally low genetic diversity in the southernmost Cape buffalo was European colonization in South Africa, with some possible additional impact caused by the rinderpest epidemic. Both populations had population sizes below 100 animals during

the last century and were isolated from other populations (de Jager et al., 2021; O'Ryan et al., 1998), preventing the introduction of genetic variants by immigration, which may have contributed to the rebound of Cape buffalo populations elsewhere (Smits et al., 2013, 2014; Van Hooft et al., 2000).

In terms of conservation implications, the overall genetic diversity in Cape buffalo is very high compared to many other large mammal species (de Jager et al., 2021) (Figure 2), and the genetic diversity in the most genetically eroded populations is still not alarmingly low at present. However, there is still cause for concern regarding the extensive population declines, recent loss of heterozygosity and increased inbreeding in the southernmost populations. The viability of a population is not well measured by present heterozygosity alone, because the temporal trajectory of population size also significantly contributes to its fitness and genetic load (Kyriazis et al., 2021). In fact, one of the most concerning trajectories is the exact scenario playing out in the two southernmost Cape buffalo populations, in which a historically large population experiences a rapid and severe reduction in numbers. This means that the population will not have had time to purge recessive deleterious variants, as seen in the critically endangered kakapo (Dussex et al., 2021). Furthermore, it has been shown that in red deer (*Cervus elaphus*) even moderate levels of inbreeding across the genome ($F_{\text{genomic}} = 0.125$) can have significant effects on adult fitness traits, such as lifetime breeding success (LBS), reducing female LBS by 72% and male LBS by 95% (Huisman et al., 2016). In addition, our results suggest historical connectivity even across large continental distances (e.g., gene flow from northern to southern populations), which may have contributed to maintaining high genetic diversity. Such connectivity is certainly threatened by ongoing human-activity-induced habitat fragmentation (Ernest et al., 2012). We therefore highlight that Cape buffalo populations may be particularly sensitive to rapid reductions in genetic diversity, especially when this cannot be recovered by subsequent gene flow due to such habitat fragmentation and confinement to isolated protected areas (Heller et al., 2010). The two southernmost populations may be portents of the genetic consequences of unabated habitat loss and fragmentation across sub-Saharan Africa. Therefore, monitoring of signs of reduced fitness in such populations is warranted.

This finding illustrates the value of studies linking demographic history—including very recent history—to patterns of genetic diversity in order to identify conservation challenges in vulnerable species.

In summary, we have expanded the set of available Cape buffalo samples to much more broadly cover the Cape buffalo range as opposed to only the southern tip of the range. This, combined with the use of a whole genome sequencing approach focused not only on genetic diversity levels but also with associated causative factors, has allowed us to identify two populations that have markedly reduced genetic diversity and to uncover that this was due to a rapid decline driven by colonialism. Importantly, the fact that it was due to a rapid decline means that these populations may be in more particular need of management effort. This further illustrates the value of our approach in identifying vulnerable populations.

FUNDING INFORMATION

L.Q. and I.M. were funded by the European Research Council (ERC-2018-STG-804679). A.B.-O. was supported by a Carlsberg Foundation Reintegration Fellowship (grant no. CF19-0427). M.-H.S.S. was supported by a Carlsberg Foundation Reintegration Fellowship (grant no. CF20-0355). U.S. Department of Agriculture, Agricultural Research Service appropriated projects 5438-32000-034-00D (M.P.H.) and 3040-31000-100-00D (T.P.L.S.). D.d.J. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101026951. H.R.S. was funded by Independent Research Fund Denmark (8021-00344B).

AUTHOR CONTRIBUTIONS

L.Q., A.A., R.H. and I.M. conceived and designed the study. L.Q., G.G.-E., C.S., A.B.-O., X.L., P.P., L.D.B., K.H., M.S.R. analyzed the data. M.-H.S.S., M.P.H. and T.P.L.S. provided samples and the context for the samples needed in the study. L.Q. wrote the initial draft of the manuscript. L.Q., A.A., R.H. and I.M. finalized the manuscript with input from all authors.

ACKNOWLEDGEMENTS

The authors thank Amal Al-Chaer for her laboratory work related to the study.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

FASTQ files containing raw read data and associated metadata for all new samples used in this study are now available for download at the European Nucleotide Archive under project accession code PRJEB58638 and submission code ERA19584828.

ORCID

Liam Quinn  <https://orcid.org/0000-0002-3597-2948>
 Genís Garcia-Erill  <https://orcid.org/0000-0003-3150-1708>
 Xiaodong Liu  <https://orcid.org/0000-0001-8839-2659>
 Mikkel-Holger S. Sinding  <https://orcid.org/0000-0003-1371-219X>
 Michael P. Heaton  <https://orcid.org/0000-0003-1386-1208>
 Deon de Jager  <https://orcid.org/0000-0003-4388-2624>
 Rasmus Heller  <https://orcid.org/0000-0001-6583-6923>
 Ida Moltke  <https://orcid.org/0000-0001-7052-8554>

REFERENCES

- Albrechtsen, A., Korneliusen, T. S., Moltke, I., van Overseem Hansen, T., Nielsen, F. C., & Nielsen, R. (2009). Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genetic Epidemiology*, 33(3), 266–274.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.
- Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research*, 23(9), 1514–1521. <https://doi.org/10.1101/gr.154831.113>
- Boitard, S., Rodríguez, W., Jay, F., Mona, S., & Austerlitz, F. (2016). Inferring population size history from large samples of genome-wide molecular data - an approximate Bayesian computation approach. *PLoS Genetics*, 12(3), e1005877.
- Browning, B. L., & Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *The American Journal of Human Genetics*, 88, 173–182. <https://doi.org/10.1016/j.ajhg.2011.01.010>
- Brüniche-Olsen, A., Kellner, K. F., Belant, J. L., & DeWoody, J. A. (2021). Life-history traits and habitat availability shape genomic diversity in birds: Implications for conservation. *Proceedings of the Biological Sciences/the Royal Society*, 288(1961), 20211441.
- Carruthers, J. (2005). Changing perspectives on wildlife in southern Africa. C.1840 to C.1914. *Society & Animals*, 13, 183–200. <https://doi.org/10.1163/1568530054927762>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- D'Amato, M. E., Alechine, E., Cloete, K. W., Davison, S., & Corach, D. (2013). Where is the game? Wild meat products authentication in South Africa: A case study. *Investigative Genetics*, 10. <https://doi.org/10.1186/2041-2223-4-6>
- de Jager, D., 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. <https://doi.org/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.
- DeCandia, A. L., Schrom, E. C., Brandell, E. E., Stahler, D. R., & vonHoldt, B. M. (2021). Sarcoptic mange severity is associated with reduced genomic variation and evidence of selection in Yellowstone National Park Wolves (*Canis lupus*). *Evolutionary Applications*, 14, 429–445. <https://doi.org/10.1111/eva.13127>
- Dussex, N., der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korch, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., ... Dalén, L. (2021). Population genomics of the critically endangered Kākāpō. *Cell Genomics*, 1(1), 100002.
- Ernest, E. M., Haanes, H., Bitanyi, S., Fyumagwa, R. D., Msoffe, P. L., Bjørnstad, G., & Røed, K. H. (2012). Influence of habitat fragmentation on the genetic structure of large mammals: Evidence for increased structuring of African Buffalo (*Syncerus caffer*) within the Serengeti ecosystem. *Conservation Genetics*, 13, 381–391. <https://doi.org/10.1007/s10592-011-0291-0>
- Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131–140.
- Fu, P.-C., Gao, Q.-B., Zhang, F.-Q., Xing, R., Wang, J.-L., Liu, H.-R., & Chen, S.-L. (2016). Gene flow results in high genetic similarity between (Rosaceae) species in the Qinghai-Tibetan plateau. *Frontiers in Plant Science*, 7(October), 1596.
- Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26(20), 5369–5406.
- García-Erill, G., & Albrechtsen, A. (2020). Evaluation of model fit of inferred admixture proportions. *Molecular Ecology Resources*, 20(4), 936–949. <https://doi.org/10.1111/1755-0998.13171>
- García-Erill, G., Jørgensen, C. H. F., Muwanika, V. B., Wang, X., Rasmussen, M. S., de Jong, Y. A., Gaubert, P., Olayemi, A., Salmona, J., Butynski, T. M., Bertola, L. D., Siegmund, H. R.,

- Albrechtsen, A., & Heller, R. (2022). Warthog genomes resolve an evolutionary conundrum and reveal introgression of disease resistance genes. *Molecular Biology and Evolution*, 39(7). <https://doi.org/10.1093/molbev/msac134>
- Goodsman, D. W., Cooke, B., Coltman, D. W., & Lewis, M. A. (2014). The genetic signature of rapid range expansions: How dispersal, growth and invasion speed impact heterozygosity and allele surfing. *Theoretical Population Biology*, 98(December), 1–10.
- Gopalan, S., Smith, S. P., Korunes, K., Hamid, I., Ramachandran, S., & Goldberg, A. (2022). Human genetic admixture through the lens of population genomics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1852). <https://doi.org/10.1098/rstb.2020.0410>
- Heaton, M. P., Keele, J. W., Harhay, G. P., Richt, J. A., Koohmaraie, M., Wheeler, T. L., Shackelford, S. D., Casas, E., King, D. A., Sonstegard, T. S., van Tassell, C. P., Neibergs, H. L., Chase, C. C., Jr., Kalbfleisch, T. S., Smith, T. P. L., Clawson, M. L., & Laegreid, W. W. (2008). Prevalence of the prion protein gene E211K variant in U.S. cattle. *BMC Veterinary Research*, 4(July), 25.
- Heller, R., Okello, J. B. A., & Siegmund, 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.
- Heller, R., Brüniche-Olsen, A., & Siegmund, H. R. (2012). Cape Buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. *Molecular Ecology*, 21(16), 3947–3959.
- Hempel, E., Faysal Bibi, J., Faith, T., Brink, J. S., Kalthoff, D. C., Kamminga, P., Paijmans, J. L. A., Westbury, M. V., Hofreiter, M., & Zachos, F. E. (2021). Identifying the true number of specimens of the extinct blue antelope (*Hippotragus leucophaeus*). *Scientific Reports*, 11(1), 2100.
- Huisman, J., Kruuk, L. E. B., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, 113(13), 3585–3590. <https://doi.org/10.1073/pnas.1518046113>
- IUCN. (2018). *Syncerus caffer*: IUCN SSC antelope specialist group. *IUCN Red List of Threatened Species*. <https://doi.org/10.2305/iucn.uk.2019-1.rlts.t21251a50195031.en>
- Jiang, Z., Hubley, R., Smit, A., & Eichler, E. E. (2008). DupMasker: A tool for annotating primate segmental duplications. *Genome Research*, 18, 1362–1368. <https://doi.org/10.1101/gr.078477.108>
- Kalbfleisch, T. S., Murdoch, B. M., Smith, T. P. L., Murdoch, J. D., Heaton, M. P., & McKay, S. D. (2018). A SNP resource for studying north American moose. *F1000Research*, 7(January), 40.
- Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Chris Funk, W. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 118(48). <https://doi.org/10.1073/pnas.2104642118>
- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(November), 356.
- Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, 5(1), 33–47.
- Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-05257-7>
- Leonard, J. A., Rohland, N., Glaberman, S., Fleischer, R. C., Caccione, A., & Hofreiter, M. (2005). A rapid loss of stripes: The evolutionary history of the extinct quagga. *Biology Letters*, 1(3), 291–295.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].
- Li, H., Xiang-Yu, J., Dai, G., Zhili, G., Ming, C., Yang, Z., Ryder, O. A., Li, W.-H., Yun-Xin, F., & Zhang, Y.-P. (2016). Large numbers of vertebrates began rapid population decline in the late 19th century. *Proceedings of the National Academy of Sciences of the United States of America*, 113(49), 14079–14084.
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496.
- Lorenzen, E. D., Heller, R., & Siegmund, H. R. (2012). Comparative Phylogeography of African Savannah ungulates. *Molecular Ecology*, 21(15), 3656–3670.
- Maier, R., Flegontov, P., Flegontova, O., Changmai, P., & Reich, D. (2022). On the limits of fitting complex models of population history to genetic data. *bioRxiv*. <https://doi.org/10.1101/2022.05.08.491072>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - fast D-statistics and related admixture Evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, 210(2), 719–731.
- Metzger, K. L., Sinclair, A. R. E., Ray Hilborn, J., Hopcraft, G. C., & Mduma, S. A. R. (2010). Evaluating the protection of wildlife in parks: The case of African Buffalo in Serengeti. *Biodiversity and Conservation*, 19, 3431–3444. <https://doi.org/10.1007/s10531-010-9904-z>
- Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21(1), 94.
- Moodley, Y., Russo, I.-R. M., Dalton, D. L., Kotzé, A., Muya, S., Haubensak, P., Bálint, B., Munimanda, G. K., Deimel, C., Setzer, A., Dicks, K., Herzig-Straschil, B., Kalthoff, D. C., Siegmund, H. R., Robovský, J., O'Donoghue, P., & Bruford, M. W. (2017). Extinctions, genetic erosion and conservation options for the black rhinoceros (*Diceros bicornis*). *Scientific Reports*, 7(February), 41417.
- Morin, P. A., Archer, F. I., Avila, C. D., Balacco, J. R., Bukhman, Y. V., Chow, W., Fedrigo, O., Formentí, G., Fronczek, J. A., Functammasan, A., Gulland, F. M. D., Haase, B., Peter Heide-Jorgensen, M., Houck, M. L., Howe, K., Misuraca, A. C., Mountcastle, J., Musser, W., Paez, S., ... Jarvis, E. D. (2021). Reference genome and demographic history of the Most endangered marine mammal, the Vaquita. *Molecular Ecology Resources*, 21(4), 1008–1020.
- Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., Demboski, J. R., Doll, A., da Fonseca, R. R., Fulton, T. L., Gilbert, M. T. P., Heintzman, P. D., Letts, B., McIntosh, G., O'Connell, B. L., Peck, M., Pipes, M.-L., Rice, E. S., Santos, K. M., ... Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, 358(6365), 951–954.
- Ndow, G., Radeino Ambe, J., & Tomori, O. (2019). Emerging infectious diseases: A historical and scientific review. In *Socio-cultural dimensions of emerging infectious diseases in Africa* (pp. 31–40). Springer International Publishing.
- NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors (2018). *BMC Bioinformatics*, 19(1):536. <https://doi.org/10.1186/s12859-018-2579-2>
- Njeumi, F., Taylor, W. D., Diallo, A., Miyagishima, K., Pastoret, P.-P., Vallat, B., & Traore, M. (2012). The long journey: A brief review of the eradication of rinderpest. *Revue Scientifique et Technique de l'OIE*, 31, 729–746. <https://doi.org/10.20506/rst.31.3.2157>
- Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson,

- P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliusen, T., Malaspina, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., ... Willerslev, E. (2013). Recalibrating equus evolution using the genome sequence of an early middle Pleistocene horse. *Nature*, 499, 74–78. <https://doi.org/10.1038/nature12323>
- 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, 85–94. <https://doi.org/10.1111/j.1469-1795.1998.tb00015.x>
- Pečnerová, P., Garcia-Erill, G., Liu, X., Nursyifa, C., Waples, R. K., Santander, C. G., Quinn, L., Frandsen, P., Meisner, J., Stæger, F. F., Rasmussen, M. S., Brüniche-Olsen, A., Hviid Friis Jørgensen, C., da Fonseca, R. R., Siegismund, H. R., Albrechtsen, A., Heller, R., Moltke, I., & Hanghøj, K. (2021). High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Current Biology*, 31(9), 1862–71.e5.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967.
- Pierskalla, J., De Juan, A., & Montgomery, M. (2019). The territorial expansion of the colonial state: Evidence from German East Africa 1890–1909. *British Journal of Political Science*, 49(2), 711–737.
- Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., Veeramah, K. R., Woerner, A. E., O'Connor, T. D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A. E., Malig, M., Hernandez-Rodriguez, J., ... Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, 499(7459), 471–475.
- Rashid, N. (2014). British colonialism in East-Africa during nineteenth century. *IOSR Journal of Humanities and Social Science*, 19, 8–11. <https://doi.org/10.9790/0837-19310811>
- Saremi, N. F., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalén, L., Figueiró, H. V., Johnson, W. E., Milne, H. J., O'Brien, S. J., O'Connell, B., Onorato, D. P., Riley, S. P. D., Sikich, J. A., Stahler, D. R., Vilella, P. M. S., Vollmers, C., Wayne, R. K., Eizirik, E., ... Shapiro, B. (2019). Puma genomes from north and South America provide insights into the genomic consequences of inbreeding. *Nature Communications*, 10. <https://doi.org/10.1038/s41467-019-12741-1>
- Schraiber, J. G., & Akey, J. M. (2015). Methods and models for unravelling human evolutionary history. *Nature Reviews. Genetics*, 16(12), 727–740.
- Sjödin, P., McKenna, J., & Jakobsson, M. (2021). Estimating divergence times from DNA sequences. *Genetics*, 217(4). <https://doi.org/10.1093/genetics/iyab008>
- Smits, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., & Caron, A. (2013). Pan-African genetic structure in the African Buffalo (*Syncerus caffer*): Investigating intraspecific divergence. *PLoS One*, 8(2), e56235.
- Smits, N., Cornélis, D., Chardonnet, P., Caron, A., de Garine-Wichatitsky, M., Jori, F., Mouton, A., Latinne, A., Pigneur, L.-M., Melletti, M., Kanapeckas, K. L., Marescaux, J., Pereira, C. L., & Michaux, J. (2014). Genetic structure of fragmented southern populations of African Cape Buffalo (*Syncerus caffer caffer*). *BMC Evolutionary Biology*, 14(November), 203.
- Summers, C. (1994). *From civilization to segregation: Social ideals and social control in southern Rhodesia, 1890–1934*. Ohio University Press.
- Tizard, I. R. (2021). A brief history of veterinary vaccines. *Vaccines for Veterinarians*, 1–12. <https://doi.org/10.1016/b978-0-323-68299-2.00010-1>
- Van Hooft, W. F., Groen, A. F., & Prins, H. H. (2000). Microsatellite analysis of genetic diversity in African Buffalo (*Syncerus caffer*) populations throughout Africa. *Molecular Ecology*, 9(12), 2017–2025.
- 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.
- Van Sittert, L. (2005). Bringing in the wild: The commodification of Wild animals in the cape colony/province C. 1850–1950. *Journal of African History*, 46(2), 269–291.
- Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, 28, 35–48. <https://doi.org/10.1111/mec.14954>
- Weng, Z.-Q., Saatchi, M., Schnabel, R. D., Taylor, J. F., & Garrick, D. J. (2014). Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genetics, Selection, Evolution*, 46(May), 34.
- Young, E. (1969). The significance of infectious diseases in African game populations. *Zoologica Africana*, 4(2), 275–281.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Quinn, L., Garcia-Erill, G., Santander, C., Brüniche-Olsen, A., Liu, X., Sinding, M.-H., Heaton, M. P., Smith, T. P. L., Pečnerová, P., Bertola, L. D., Hanghøj, K., Rasmussen, M. S., de Jager, D., Siegismund, H. R., Albrechtsen, A., Heller, R., & Moltke, I. (2023). Colonialism in South Africa leaves a lasting legacy of reduced genetic diversity in Cape buffalo. *Molecular Ecology*, 32, 1860–1874. <https://doi.org/10.1111/mec.16851>