Evolutionary history of the roan antelope across its African range

Gonçalves, Margarida; Siegismund, Hans R.; van Vuuren, Bettine Jansen; Ferrand, Nuno; Godinho, Raquel

Published in:
Journal of Biogeography

DOI:
10.1111/jbi.14241

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Evolutionary history of the roan antelope across its African range

Margarida Gonçalves1,2 | Hans R. Siegismund3 | Bettine Jansen van Vuuren4 | Nuno Ferrand1,2,4 | Raquel Godinho1,2,4

Abstract

Aim: Phylogeographic studies on savanna ungulates have extensively evaluated genetic patterns mostly related to Pleistocene climatic oscillations. We address this subject through a comprehensive assessment across the pan-African range of the roan antelope, assessing whether climatic oscillations or natural physical barriers play a predominant role in the evolutionary history of the species. We also address the spatial concordance of the genetic structure with the currently recognized subspecies.

Location: Sub-Saharan Africa.

Taxon: Roan antelope (Hippotragus equinus).

Methods: We genotyped 43 microsatellite loci and sequenced whole-mitochondrial genomes for 131 individuals across the species’ entire African range. We performed spatial analyses of genetic diversity for contemporary and historical samples and calculated overall patterns of genetic differentiation and structure for both marker types. We also estimated the timing of divergence events and demographic trends, correlating these with the species’ biotic attributes as well as abiotic features shaping African savanna habitats.

Results: Our analyses uncovered highly structured clusters and contact zones across the distribution of the roan antelope, including five nuclear groups and four mitochondrial lineages. The north-west group had the highest level of intra-group diversity as well as inter-group divergence and represents the oldest vicariant event. The central and southern groups had the lowest intra-group diversity with low divergence values separating them, suggesting a more recent ancestry for these groups. All population groups showed signals of demographic stability over time but with a demographic decline during the Holocene.

Main conclusions: The roan antelope exhibits significant population structure across its African range. This structure is largely associated with natural physical barriers, whereas contact zones could more easily be explained by climatic events. Based on estimates of genetic diversity, we propose a West African ancestry for this species with subsequent eastward and southern range expansions, as well as the persistence of...
1 | INTRODUCTION

The African savanna biome, including woodlands, shrublands and grasslands, extends across much of the sub-Saharan region (Mayaux et al., 2004) and is home to the greatest diversity of ungulates on earth (including rhinoceros, giraffes, and antelopes) (Du Toit & Cumming, 1999). The emergence of much of this spectacular diversity is dated to around 2.8 million years ago and coincides with the increased dominance of grasslands due to a climatic shift towards cooler temperatures and increased aridity (Dupont, 2011; Vrba, 1996). Specifically, a notable contributing factor that drove much of the speciation events is habitat fluctuations during Pleistocene climatic cycles (DeMenocal, 2004; Lorenzen et al., 2012), combined with continental-wide geomorphological changes throughout the Pliocene and Pleistocene (Badgley, 2010; Goudie, 2005; Trauth et al., 2009).

In sub-Saharan Africa, the Pleistocene was marked by alternating cycles of cool and dry interpluvials and warm and wet pluvials. These cycles drove repeated expansions and contractions of savanna habitats, and consequently led to the expansion and fragmentation of savanna-adapted species (DeMenocal, 2004). The Pleistocene refuge theory postulates that the maintenance of grassland refugia during pluvials enabled the persistence of species through repeated ice ages, promoting the divergence between populations in different refugia during periods of isolation, as well as the accumulation of genetic diversity because of populations’ persistence. During interpluvials, recolonization of newly available habitats from these source populations allowed for the spread of diversity (reviewed in Hewitt, 2004). Range expansion rates during interpluvials and the presence of geomorphological barriers (such as rivers and mountains) dictated the position of contact zones, and ultimately shaped the spatial patterns that we see today.

Several studies have documented spatial genetic patterns among African ungulates (see Arctander et al., 1999; Lorenzen et al., 2010; Smits et al., 2013), mostly relating results to Pleistocene climatic oscillations. Based on their own work and data available for other species, Lorenzen et al. (2012) suggested refugia in Southern and West Africa, with East Africa harbouring a melting pot of diversity and a mosaic of smaller refugia. The complex patterns in East Africa were largely driven by high species richness and turnover, exacerbated by habitat instability associated with the continued volcanic and tectonic activity of the East African Rift System (EARS) (Moorley & Kingdon, 2013; Trauth et al., 2009). However, for species with pan-African ranges, spatial genetic patterns remain poorly understood, likely due to the critical sampling effort required to cover whole species’ ranges and, hence, the paucity of studies that include such data.

The roan antelope (*Hippotragus equinus*) is the second-largest antelope in Africa and one of the two extant species within this genus. It has a wide distribution throughout sub-Saharan savanna habitats, occurring from West towards East and Southern Africa (Chardonnet & Crosmary, 2013; IUCN SSC Antelope Specialist Group, 2017). Lorenzen et al. (2012) included data for this species in their comparative assessment of African antelope’s spatial patterns, but the data available at the time was incomplete and covered only sections of the pan-African range of the roan antelope (Alpers et al., 2004; Matthee & Robinson, 1999). In addition, the spatial genetic patterns that emerged from these studies challenged the validity of the recognized subspecies (Ansell, 1972). Instead, it led to the proposal for one Evolutionary Significant Unit situated in West Africa and referable to the subspecies *H. e. koba*, whereas the remaining range constituted a geographically more diverged group whose genetic structuring did not strictly agree with subspecies boundaries, namely for *H. e. lang-heldi* in East Africa, *H. e. cottoni* in south-central Africa, and *H. e. equinus* in the southern Africa. No or very limited sampling was available for *H. e. charicus* in north-central and *H. e. bakeri* in north-east Africa. Studies that suffer from incomplete geographic sampling may lead to partial or even erroneous understanding of spatial genetic patterns and hence, also to inaccurate interpretations of the biogeographic processes driving these patterns.

Biologically inherent characteristics may also influence the evolutionary history and demographic trends of species. Although having a continental range, roan antelopes have a patchy and discontinuous distribution at the local scale due to specific habitat requirements that include mesic savanna woodlands and the constant availability of standing water. They are gregarious animals with a strong social structure, and breeding herds consist of one dominant male and up to 15 females and calves (Chardonnet & Crosmary, 2013). Despite being listed as ‘Least Concern’ by the IUCN (IUCN SSC Antelope Specialist Group, 2017), most wild populations only occur within protected areas; it is acknowledged that these animals are more common throughout West and Central Africa, whereas across some parts of its eastern and southern distribution, populations are considered rare and even virtually extinct (Chardonnet & Crosmary, 2013; East & IUCN SSC Antelope Specialist Group, 1999).

Overall, the roan antelope is an ideal model for a comprehensive biogeographic overview of savanna ungulates with pan-African ranges. Based on previous work for the species (Alpers et al., 2004; Matthee & Robinson, 1999) and on the climatic and
geomorphic history of sub-Saharan Africa, we hypothesize that the roan antelope harbours marked genetic signatures primarily concordant with natural physical barriers (mountains, rivers and forest), despite Pleistocene climatic oscillations having established local contact zones and possibly mitochondrial introgressions. Grounded on the patchy distribution and different geographical abundances of the species (Chardonnet & Crosmary, 2013; East & IUCN SSC Antelope Specialist Group, 1999), we further hypothesize that current roan antelope populations exhibit uneven genetic diversity, as previously evaluated using a much smaller number of nuclear markers and a small fragment of mitochondrial DNA (Alpers et al., 2004), and have genetic signatures of demographic declines in the recent past. To test these hypotheses, we provide the most comprehensive phylogeographic assessment of this species to date, based on contemporary and historical samples from across its entire range, and on a large set of microsatellite markers and complete mitochondrial genomes. Our results give a genetic context for the recognized subspecies, disclose contact zones and allow us to provide suggestions for this iconic species’ management and conservation. We address the evolutionary history of the roan antelope against the backdrop of the climatic and geomorphological history of the African continent throughout the Pleistocene, and discuss the importance of geographical barriers and their permeability for large ungulates with a pan-African distribution.

2 MATERIALS AND METHODS

2.1 Sample collection and DNA extraction

Our dataset included 61 contemporary (1993–2004) and 70 historical (1842–1973) samples from 70 sampling localities, representing the entire natural range of the roan antelope, including all six subspecies described by Ansell (1972) (Table S1). For the contemporary samples, total genomic DNA was extracted from muscle samples using the Qiagen DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer’s instructions. DNA extractions for the historical samples (skin, skull, bone, and tooth) followed Dabney et al. (2013) in a laboratory dedicated to low-quality DNA work (see Text S1 for details). For both methods, blank controls were routinely included while duplicate (for contemporary) and triplicate (for historic) samples were amplified to minimize possible amplification errors. Using GENEMAPPER 4.0 (Applied Biosystems), allele sizes were scored against the GeneScan 500 LIZ Size Standard and manually verified. Consensus genotypes were built from the replicates, and only samples with <25% missing data were included in subsequent analyses. Pairwise relatedness was calculated for individuals from the same locality using the Ritland (1996) estimator available in GENEALEX 6.5.3 (Peakall & Smouse, 2006, 2012). Pairwise relatedness values above 0.25 are indicative of high relatedness, suggesting that individuals belong to the same family group (such as parent-offspring, full and/or half-siblings); the inclusion of highly related individuals can bias downstream analyses of population structure, especially when analyses are based on allele frequencies (see O’Connell et al., 2019). Therefore, for each pairwise relatedness value >0.25, one individual was randomly selected and removed from further analyses. This procedure led to the removal of four individuals from our dataset.

2.3 Nuclear data analyses

Spatial genetic structure (SGS) was first assessed for the whole dataset using STRUCTURE 2.3.4 (Falush et al., 2003; Hubisz et al., 2009; Pritchard et al., 2000); we selected the admixture model with correlated allele frequencies. Analyses were performed for K = 1 to K = 10 across 10 independent runs per K, with 1 million Markov Chain Monte Carlo (MCMC) iterations following a burn-in period of 0.5 million steps. STRUCTURE HARVESTER WEB 0.6.94 (Earl & VonHoldt, 2012) was used to plot the likelihood for each K using the mean likelihood and delta-K statistics (Evanno et al., 2005). Results from this first analysis (run 1) revealed the highest delta-K value for K = 2 (Figure S1b). However, the evaluation of highest delta-K alone might not be reliable under complex evolutionary scenarios (Janes et al., 2017). Additionally, genetic substructure for subsequent values of K (Figure S1c), prompted us to partitioning data according to inferred clusters for K = 2, as suggested by Pritchard et al. (2009). Therefore, a second STRUCTURE run (run 2) was performed, following a hierarchical analysis on each of the two main clusters, named A and B, using the same parameters as for run 1 (Figure 1a). To assess whether there is a correspondence between SGS for contemporary and historical samples, a third STRUCTURE analysis (run 3) was performed for these two datasets separately. Our criterion to choose the best K in STRUCTURE runs was the lowest value of K in which the L(K) curvature starts to plateau. Lastly, we also tested for the presence of distinct genetic clusters across the entire dataset based on Nei’s genetic distances through principal coordinate analyses (PCoA) in GENEALEX. ARLEQUIN 3.5 (Excoffier & Lischer, 2010) was used to test for deviations from Hardy–Weinberg equilibrium (HWE) within each group that resulted from the STRUCTURE analyses, as well as linkage disequilibrium (LD) between pairs of loci. The same groups were used in ARLEQUIN to calculate summary statistics within groups including
FIGURE 1  Nuclear microsatellite data analyses for the roan antelope. (a) Bayesian clustering analysis for $K = 2$, using the complete dataset (run 1); (b) Bayesian hierarchical clustering analyses for $K = 1$ (cluster A) and $K = 4$ (cluster B). For (a, b), graphs show the individual assignment proportions ($0.0 < q < 1.0$), according to STRUCTURE results. Individuals are represented by a vertical bar, and each colour symbolizes a different nuclear group: north-west in blue; north-east in green; eastern in yellow; central in orange and southern in red. Individuals ordered according to sampling-site, as in Table S1. Triangles indicate historic samples. Squares represent the mitochondrial lineage for each individual: north-west in blue; north-east in green; east-central in yellow and southern in red. White squares represent individuals with no mtDNA information. (c) Geographic distribution of nuclear groups, according to the five main groups identified by STRUCTURE for the hierarchical clustering analyses. Pie charts are plotted according to sampling locality, indicated by numbers (from 1 to 70, as in Table S1), and represent the average assignment probability to each cluster of individuals sampled within each locality. Species distribution retrieved from IUCN SSC Antelope Specialist Group (2017) and subspecies limits adapted from Ansell (1972). Also depicted are geographic features mentioned in the text, namely the East African Rift System (EARS) in black, and the Zambezi River in dark blue. (d) Scatterplot of the first two principal coordinates from principal coordinate analyses results, using the complete dataset. Colours correspond to the main five groups identified by the hierarchical clustering analyses. Each point represents an individual. Circles depict contemporary samples, whereas triangles represent historic samples.
expected heterozygosity \( (H_e) \), the mean number of alleles per locus \( (N_a) \), number of private alleles \( (N_{pa}) \) and inbreeding coefficient \( (F_{IS}) \). Allelic richness (AR) was estimated using FSTAT 2.9.3 (Goudet, 2001).

All summary statistics were calculated for the full dataset as well as for contemporary and historical samples separately. Following Bonferroni corrections for multiple comparisons, significance levels were assessed against a \( p \)-value of 0.01 (Rice, 1989). Levels of differentiation between groups were calculated using the pairwise means of the fixation index \( (F_{ST}) \) and the standardized \( C_1^{ST} \) statistic (Hedrick, 2005) \( (C_1^{ST}) \) corrects \( F_{ST} \) biases for microsatellite loci. We also performed an analysis of molecular variance (AMOVA) in ARLEQUIN with 1000 permutations and allowing for 5% missing data per locus.

Historical variation in the effective population size \( (N_e) \) for each group was inferred using the full dataset and the package ‘vareff’ (Nikolic & Chevalet, 2014) implemented in R 3.4.1 (R Core Team, 2017). This method uses a coalescent framework to estimate approximate likelihoods of the distribution of allele frequencies. Demographic parameters were modelled following the authors’ recommendations. Variation in the number of demographic changes (the JMAX parameter) led to only slight changes in the steepness of the curves. Two mutation rates were assessed: \( 1 \times 10^{-3} \) and \( 1 \times 10^{-4} \) mutations/locus/year (Qi et al., 2015; Ritz et al., 2000) for both single- and two-phase mutation models, with 20% motif steps in the latter. Models were assessed by inspecting the posterior distribution of trial runs, using 10 million steps MCMC chains, with a 10% burn-in. Best runs within groups were plotted using the NTDist function in R. Time estimates were converted to years using a generation time of 8.4 years (Pacifici et al., 2013), as well as 6 years, which derives from an ecological assessment of the age at which males mature (Chardonnet & Crosmary, 2013).

2.4 Whole mitochondrial sequencing and assembly

For contemporary samples, double-stranded DNA libraries were prepared using Nextera XT DNA Library Prep Kit (Illumina), following the modifications described in Tan and Mikhailiev (2016). After a sonication step, size selection of DNA fragments of ca. 350 bp was performed using AMPure XP bead clean-up (Beckman Coulter). For historical samples, DNA library preparation followed Meyer and Kircher (2010) with modifications proposed by Kircher et al. (2012). Amplification and purification of these DNA libraries followed Dabney and Meyer (2012). We performed a target sequencing capture, using four fragments of ca. 5000 bps as bait, covering the entire mitochondrial genome of the roan antelope. Bait production and capture reactions were performed as described in Rocha (2014) using the same primers. Modifications during long-range PCR amplifications were done to improve the specificity for the roan antelope (see Text S3 for details). Considering the limited availability of bait product, we prioritized capture reactions for historical samples decreasing to one-third the bait used for capture reactions of contemporary samples. Following capture, both contemporary and historical enriched libraries were quantified by Kappa Library Quantification Kit (Roche Sequencing and Life Science, USA) and pooled at equimolar amounts. Each pool, containing 42 contemporary and 70 historical samples, respectively, was sequenced on the MiSeq platform (Illumina) on a single cell using the 250-bp paired-end sequencing protocol.

Raw sequences were demultiplexed after adapter trimming using trimmomatic 0.36 (Bolger et al., 2014); only paired-end reads longer than 35 bp were retained. Sequences were mapped against an \( H. \ equinus \) reference mitogenome (NCBI NC_020712.1) using the function \( \text{mem} \) from the mapper \( \text{bwa} \) 0.7.16 (Li, 2013). Base calling was performed using samtools 1.5 and bcftools 1.9 (Li et al., 2009) (mapping quality >30 and base coverage >3x). Quality mapped-reads for historical samples were on average 30%, with an average depth of coverage to the mitogenome of 200x. Quality mapped-reads for contemporary samples were on average 8%, with an average depth of coverage to the mitogenome of 23x. For contemporary samples, base calling was complemented using the \( \text{abyss-pe} \) 1.3.4 de novo assembler (Simpson et al., 2009) with default parameters. The resulting scaffolds were blasted against a modified circular reference genome and matching scaffolds were combined and aligned against mapped sequences to reach a consensus on dubious positions for each sample. Only positions with >5x coverage were considered, and a threshold of 25% was allowed for unassigned bases (i.e. poor quality bases \( [N] \)).

2.5 Mitochondrial data analysis

The final dataset was aligned using \( \text{muscle} \) 3.8.31 (Edgar, 2004). Coding regions were translated using \( \text{dnasp} \) 5.0 (Librado & Rozas, 2009); no non-sense mutations were detected. The alignment for the hypervariable region (HVR) was unresolved due to a high number of polymorphic sites, and the region was excluded from all analyses. The \( H. \ niger \) mitogenome (NCBI NC_020713.1) was used as an outgroup. To visualize the relationship among haplotypes, their spatial distribution, as well as haplotype sharing within and between localities, a median-joining network (Bandelt et al., 1999) was constructed using \( \text{popart} \) 1.7 (http://popart.otago.ac.nz). \( \text{arlequin} \) was used to assess mitochondrial genetic diversity within each lineage and nuclear group through both haplotype \( (H_d) \) and nucleotide \( (\pi) \) diversities, as well as mean pairwise distances, defined as the mean number of differences between all pairs of haplotypes. For these analyses, we allowed 5% missing data. Genetic diversity was also assessed per coding region across the mitogenomes. To assess the hierarchical partitioning of variation, an AMOVA with 1000 permutations was performed, both within and between groups. Summary statistics were also calculated and compared for the contemporary and historical samples separately.

To infer divergence times for the roan antelope lineages, we included eight mitogenomes available from public repositories as outgroups. Five of these mitogenomes belonged to the remaining extant species of the Hippotraginae sub-family, namely the East
analyses were performed in codon-position of the protein-coding genes. Bayesian phylogenetic (wildebeest (Connochaetes gnou), the hartebeest (Alcelaphus buselaphus; NC_020676.1), the black wildebeest (Connochaetes gnou; NC_020698.1), and the bontebok (Damaliscus pygargus; NC_020627.1). These eight mitogenomes were aligned against the roan and the sable antelope reference mitogenomes, as well as four samples from the present study representing each mitochondrial lineage of the roan antelope (see Table S1 for details). PartitionFinder 2.0 (Lanfear et al., 2016) was used to infer the optimal partitioning strategy. Corrected Akaike’s information criterion (AICc) score was used to select both the best partition scheme as well as substitution model for each partition. The software was run with linked branch-lengths, under a greedy search mode in phyml (Guindon et al., 2010; Lanfear et al., 2012). The best scheme partitioned the data into three blocks: RNA loci; first and second codon-position of the protein-coding genes; third codon-position of the protein-coding genes. Bayesian phylogenetic analyses were performed in BEAST 2.6.3 (Bouckaert et al., 2014) using the three partitions and corresponding best-fit models (GTR+I+G), under an uncorrelated relaxed log-normal molecular clock model. To avoid overparameterization, we assumed linked trees for the partitions. From Bibi’s (2013) mitochondrial phylogeny for Bovidae, which used fossils for multi-calibrated points, we retrieved calibration ages as priors for the root node and main nodes of the Hippotraginae and Alcelaphinae sub-families (Table S3). Both calibrated and non-calibrated nodes were set under a lognormal distribution, assuming monophyly. The Calibrated Yule speciation process was used to account for sequences between species. Two independent MCMC were run for 50 million generations, with 10% burn-in and logging results every 50 thousand generations. Convergence of the posterior distributions for both runs was assessed using TRACER 1.7.1 (Rambaut et al., 2018) considering only models whose statistics had an effective sample size value above 1000. Maximum clade credibility trees were estimated with TREEANNOTATOR 2.5.1, discarding a 10% burn-in. The consensus tree was visualized in Figtree 1.4.4.

To test for signatures of demographic expansion within each roan antelope lineage, we used the whole dataset and estimated neutrality tests based on Tajima’s D (Tajima, 1989) and Fu’s F_{Is} (Fu, 1997) with 1000 permutations, implemented in ARLEQUIN.

3 | RESULTS

3.1 | Nuclear data

The final dataset included 131 individuals from 70 localities across sub-Saharan Africa, genotyped for 43 autosomal microsatellites (Tables S1 and S2). Ten of the initial loci were excluded from our analyses because of poor amplification success. Locus HN3 was removed as it deviated significantly from HWE (0.05 < p < 0.01; after Bonferroni corrections), as a result of a strong heterozygote deficiency. Deviations also occurred in other loci but were not consistently observed across groups (as defined by the population structure analyses), and we retained these in subsequent analyses (Table S4). No significant deviations from LD were observed for any pair of loci across groups.

3.1.1 | Population structure

Our initial clustering analysis (run 1) retrieved K = 2 as the most likely number of groups, based on delta-K statistic (Figure S1b). The first group, with individual assignment probabilities >0.5 to the first cluster, corresponded to animals in West Africa, from the Central African Republic westwards (cluster A), whereas the second group included the remainder of the specimens (cluster B) (Figure 1a). However, delta-K statistic might not be reliable under complex evolutionary scenarios (Janes et al., 2017); whilst the PcoA also supported the same two main groups, with the first two components explaining 12% of the genetic variation (Figure 1d), results for increasing values of K clearly indicated additional genetic structure across the range of the roan antelope (Figure S1c). Therefore, we performed a hierarchical clustering analysis (run 2) for the two main clusters A and B (Figure 1b).

Based on the lowest value of K in which the L(K) curvature starts to plateau, the highest likelihood in cluster A was assigned to K = 1 (Figure S2a), with subsequent values resulting in the split of

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>H_e ± SD</th>
<th>N_a ± SD</th>
<th>N_pa</th>
<th>AR</th>
<th>F_{Is}</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-west</td>
<td>20</td>
<td>0.601 ± 0.331</td>
<td>6.5 ± 3.9</td>
<td>52</td>
<td>5.0</td>
<td>0.110</td>
</tr>
<tr>
<td>North-east</td>
<td>10</td>
<td>0.580 ± 0.305</td>
<td>4.7 ± 2.7</td>
<td>12</td>
<td>4.5</td>
<td>0.043</td>
</tr>
<tr>
<td>Eastern</td>
<td>33</td>
<td>0.528 ± 0.330</td>
<td>6.0 ± 4.2</td>
<td>15</td>
<td>4.3</td>
<td>0.018</td>
</tr>
<tr>
<td>Central</td>
<td>31</td>
<td>0.503 ± 0.327</td>
<td>5.2 ± 3.5</td>
<td>9</td>
<td>3.8</td>
<td>0.070</td>
</tr>
<tr>
<td>Southern</td>
<td>37</td>
<td>0.521 ± 0.331</td>
<td>6.4 ± 4.7</td>
<td>26</td>
<td>4.3</td>
<td>0.127***</td>
</tr>
</tbody>
</table>

Note: All parameters were estimated allowing 5% of missing data per locus. Abbreviations: AR, allelic richness; F_{Is}, inbreeding coefficient with *** p ≤ 0.001, after Bonferroni corrections; N, sample size; H_e, expected heterozygosity and respective standard deviations (SD); N_a, average number of alleles per locus and respective standard deviations (SD); N_pa, number of private alleles.
individuals at the level of the sampling locality (Figure S2c). Results for cluster B indicated $K = 4$ as the best model to explain data partitioning (Figure S3a). The first group of cluster B included individuals from Sudan towards Uganda; the second comprised of Tanzanian individuals; the third included individuals from Zambia and northern Mozambique and the fourth comprised of individuals from Zimbabwe, South-Africa, Botswana, Namibia and Angola (Figure 1b; Figure S3c). Combining all results from the clustering analyses, five spatially explicit clusters were considered to best explain the SGS for the roan antelope (Figure 1c). These were designated, from west to east and south, as: north-west, north-east, eastern, central and southern groups (Figure 1).

It is clear from our results that there are some areas across the range of the species where animals appear admixed (with membership to the two adjacent clusters). Although not limited to this explanation, this finding suggests the presence of secondary contact zones with historic, recent or ongoing gene flow between groups. These signatures of admixture are present between the north-west and north-east groups in South-Sudan (Figure 1c—sampling site 19), on the border between Tanzania and Zambia, between the eastern and central groups (Figure 1c—sampling site 38), and between the central and southern groups (Figure 1c—sampling site 50). For subsequent analyses, these admixed individuals were included in the groups for which they had assignment probability $>0.5$.

The consistency of these five groups across time was tentatively (because of non-ideal sampling schemes for temporal data) assessed through a third STRUCTURE analysis considering contemporary and historical samples separately. For both datasets, the best model explaining the data was assigned to $K = 3$ (Figures S4a and S5a, respectively). At this $K$, the clusters were mostly concordant with those retrieved from the hierarchical analyses from run 2 (Figures S4c and S5c). However, specimens from Zambia (the central group) did not represent a distinct group across contemporary samples, and the same can be observed for specimens from Tanzania (eastern group) across historical samples. We believe that the low sample size ($n = 3$ Zambian individuals for contemporary samples; $n = 4$ Tanzanian individuals for historical samples) explains such discrepancies. Considering the congruence in spatial patterns across the time scale covered by our samples, we are confident that the spatial genetic patterns described here precede our earliest samples. Based on these observations, we combined the contemporary and historical samples with confidence, thereby providing the most accurate nuclear spatial patterns for the roan antelope across its range.
3.1.2 | Diversity and differentiation

Our analyses of the whole dataset showed that the north-west group consistently exhibited the highest genetic diversity values, while the central group had the lowest (based on $H_e$, AR, and $N_{ps}$ statistics; Table 1). Genetic diversity estimates for contemporary and historical samples within groups showed a small but consistent decrease in diversity over time for the three groups with similar sampling sizes (north-west, north-east and southern; Table S5). The southern group is characterized by a significant inbreeding coefficient ($F_{is}$) for the combined dataset ($F_{is} = 0.127; p < 0.001$; Table 1) as well as for historical samples ($F_{is} = 0.165; p < 0.001$; Table S5), but not for the contemporary samples. This finding may simply reflect the substructure detected for the combined data (Figures S1c and S3c) and for historical samples (Figure S5c), resulting in a deficit of heterozygotes, the well-known Wahlund effect (Waples, 2015).

Pairwise $F_{ST}$ and $G_{ST}$ supported significant structure $(p < 0.001)$ across the range of roan antelopes (see Table S6). The average $F_{ST}$ value among groups was 0.111 (average $G_{ST} = 0.246$), with the highest values consistently found between the north-west and remaining groups ($0.138 < F_{ST} < 0.163$; $0.303 < G_{ST} < 0.360$). The lowest average values included comparisons between the eastern, central and southern clusters (average $F_{ST} = 0.061$; $G_{ST} = 0.134$), which is in line with the results of the PCoA plot (Figure 1d).

Our examination of population differentiation using AMOVA found significant partitioning of variation between groups (10.5%), compared with a lower percentage of variation (6.9%) attributed to differences within groups. As expected, most variation was accounted for by variation at the individual level (82.6%) (Table S7).

3.1.3 | Demographic history

In the absence of a species-specific mutation rate for the roan antelope, we assessed demographic trends using two mutation rates with a 10-fold difference. The fastest mutation rate $(1 \times 10^{-3}$ mutations/locus/year) with the two-step model yielded a better fit of the posterior values and lower 95% confidence intervals on estimated $N_e$. Taking this into account, all groups showed stable population numbers followed by a more recent effective population size decrease (Figure 2).

This trend of initial stability holds until 400 generations ago for the north-eastern and central groups, which translates to around 3.4 thousand years ago (kya) using a generation time of 8.4 years, and 2.4 kya assuming male maturity at 6 years. Following this, both the north-east and central groups showed a decrease in $N_e$ of 100-fold (Figure 2). The remaining groups show similar decreases in $N_e$ but time estimates for these events are more recent, placing it between 200 and 100 generations ago (or 1.7 to 0.8 kya for a generation time of 8.4 years and 1.2 to 0.6 kya for 6 years). Demographic estimates using the lowest mutation rate $(1 \times 10^{-4}$ mutations/locus/year) for microsatellites resulted in a ten-fold increase both in $N_e$ and in the estimated intervals (Figure S6). Although not changing the general population demographic trends, the use of a lowest mutation rate shifts inference towards older historical periods.

3.2 | Mitochondrial data

For the mitogenomic data, our dataset comprised 15,551 bp (excluding the HVR) for 82 roan antelope samples representative of the whole pan-African range of the species.

3.2.1 | Diversity and differentiation

The average haplotype and nucleotide diversities across the mitogenome were 0.993 and 0.0059, respectively. Haplotype diversity ranged between 0.971 for cytochrome c oxidase subunit I (cox1) to 0.387 for the NADH dehydrogenase subunit 3 gene (nad3). For nucleotide diversity, values varied between $\pi = 0.0109$ in NADH dehydrogenase subunit 4L (nad4L) and 0.0014 for the 12S rRNA subunit (Table S8). In total, 70 unique mitogenomes characterized the 82 samples, indicating high levels of genetic diversity in the species.

The median-joining network clustered the 70 unique haplotypes into four, well-supported lineages (Figure 3a). The majority of the haplotypes were unique to specific localities, with complete concordance in their assignment to the haplotype groupings between contemporary and historic samples. Spatially, we found good correspondence between the mitogenome and nuclear groups, but with two notable discrepancies (Figures 1b and 3b). A clear and sharp contact zone between the north-west and north-east mitogenome lineages was evident within the north-west nuclear group, with individuals to the east of Cameroon being characterized by a mitogenome haplotype typical of the north-east group, while both lineages are present in Cameroon (Figure 3b—sampling sites 9–13). The second discrepancy is the absence of unique lineages corresponding to an eastern or central nuclear group; rather, individuals from Tanzania, Zambia, Mozambique and parts of Zimbabwe and South-Africa shared the same mitogenomic lineage, hereafter referred to as the east-central mitochondrial lineage. Although a separate mitochondrial lineage—southern—corresponds to the southern nuclear group, there was extensive east-central haplotype introgression into the range of the southern lineage (Figure 3). Haplotypes from both lineages were present in the same localities in South Africa (Figure 3b—sampling site 54) and Angola (Figure 3b—sampling site 67). In addition, in northern Tanzania, where the east-central lineage occurs, a few individuals had a mitochondrial haplotype more typical of the north-eastern lineage (Figure 3b—sampling locality 29). There were only five shared haplotypes across our dataset, three in the east-central lineage and two in the southern lineage (Figure 3).

The north-west and north-east lineages have the highest values across all mitochondrial diversity estimates, including the number of substitution sites, and nucleotide and haplotype diversities (Table 2). A steep decrease in diversity estimates is seen from East Africa to...
Figure 3 Mitochondrial data analyses for the roan antelope. (a) Median-joining network based on mitochondrial genomes. Each lineage is identified by its name highlighted with a specific colour: north-west in blue; north-east in green; east-central in yellow and southern in red. Haplotypes are represented by a circle identified with the corresponding name. Triangles next to the haplotype name indicate those present in historic samples. Size of the circle is scaled according to the number of observations. Numbers next to the branches indicate the number of mutational steps between nodes. Missing haplotypes are represented by small black circles. Private haplotypes are represented in white, whereas shared haplotypes between sampling localities are coloured. (b) Geographical distribution of haplotypes. Each circle corresponds to a sampling locality (numbered in black, from 1 to 70, as in Table S1) and characterizes the haplotypes observed within that locality. Haplotypes outlined according to colours used to each lineage in (a). Shared haplotypes between localities are coloured, according to (a). Species distribution retrieved from IUCN SSC Antelope Specialist Group (2017) and subspecies limits adapted from Ansell (1972). Also depicted are geographic features mentioned in the text, namely the East African Rift System (EARS) in black and the Zambezi River in dark blue.
Southern Africa, with the lowest values registered for the southern lineage (Table 2). Evaluation of mitochondrial diversity within nuclear groups retrieved a similar pattern as for the mitogenomic lineages, but with the central group as least diverse (Table 2). AMOVA, based on the mitochondrial data, indicated that 27% of the variation is explained by differences between the groups (p < 0.0001) (Table S7). We could not perform similar analyses based on the temporal data (contemporary and historical data) as our sample sizes did not allow for meaningful comparisons (Table S9).

### 3.2.2 Divergence time estimates and demography

The inferred rate of change for the three partitions analysed varied between $2.3 \times 10^{-8}$ mutations/site/year for the third codon position of the protein-coding genes (95% higher posterior density [HPD]: 2.1–2.5 $\times 10^{-8}$) to $4.2 \times 10^{-9}$ (95% HPD: 3.6–4.7 $\times 10^{-9}$) and $5.1 \times 10^{-9}$ (95% HPD: 4.8–5.8 $\times 10^{-9}$) mutations/site/year for the first and second codon position and rRNA genes, respectively.

Bayesian analysis retrieved a highly supported tree with confidence node intervals (posterior = 1) and corresponding divergence time estimates with little overlap (Figure 4). Using Bibi’s (2013) calibration points, the tree node (common ancestor) and division between the Hippotraginae and Alcelaphinae sub-families were dated to mid- and late-Miocene, respectively (between 6.8 and 10.6 Mya). Divergence time estimates for H. potatorus sp. was dated to late-Miocene, around 5.8 Mya (95% HPD: 5.2–6.3 Mya). The main divergence events within the roan antelope were more recent and are dated to between mid- to late-Pleistocene. The oldest divergence event occurred around 591 kya (95% HPD: 452–747 kya), separating the common ancestor of the north-west lineage from the remaining ones. Around 339 kya (95% HPD: 251–436 kya), the second vicariant event separated the common ancestor of the north-east lineage, and at around 238 kya (95% HPD: 166–315 kya) the third event separated the common ancestor of east-central and southern lineages. The roan reference mitogenome (from an animal sampled from Malawi) clusters with the southern lineage, and last shared a common ancestor with the southern lineage 52 kya (95% HPD: 25–84 kya).

Neutrality tests across the mitochondrial data, retrieved negative albeit non-significant values across all lineages (Table S10), indicating no departure from the null hypothesis of demographic stability.

### 4 DISCUSSION

Roan antelopes have a wide distribution across sub-Saharan African savannas. By combining a highly comprehensive sampling scheme with a dataset of mitogenomes and nuclear microsatellites, we provide robust evidence for several distinct population groups and mitochondrial lineages, and their divergence times. We discuss the correspondence between our genetic groups and the subspecies proposed by Ansell (1972) and argue that these distinct entities came about through an interplay between biological attributes, natural physical barriers and past climatic events that shaped African biodiversity during the Pleistocene.

### 4.1 Biogeography, abiotic drivers and contact zones

Three major African natural barriers have been associated with differentiation in savanna ungulates (Grobler et al., 2005; Lorenzen et al., 2012; Vaz Pinto, 2018) as well as other mammal species (Bertola et al., 2016; McDonough et al., 2015; Visser et al., 2019). The Tropical Forest Belt spans from West Africa eastward to the EARS, and has been invoked as a major barrier for savanna-adapted terrestrial species. Large ungulates like hartebeest and the common

<table>
<thead>
<tr>
<th>TABLE 2 Measures of genetic diversity within each roan mitochondrial lineage (top) and nuclear group (bottom), based on whole mitochondrial DNA sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Mitochondrial lineages</strong></td>
</tr>
<tr>
<td>North-west</td>
</tr>
<tr>
<td>North-east</td>
</tr>
<tr>
<td>East-central</td>
</tr>
<tr>
<td>Southern</td>
</tr>
<tr>
<td><strong>Nuclear groups</strong></td>
</tr>
<tr>
<td>North-west</td>
</tr>
<tr>
<td>North-east</td>
</tr>
<tr>
<td>Eastern</td>
</tr>
<tr>
<td>Central</td>
</tr>
<tr>
<td>Southern</td>
</tr>
</tbody>
</table>

Note: All parameters were estimated allowing 5% of missing data. $\pi$, nucleotide diversity and respective standard deviation (SD); h, number of haplotypes; $H_p$, haplotype diversity and respective standard deviation (SD); MPD, mean pairwise distance and respective standard deviation (SD); N, number of samples; S, number of polymorphic (segregating) sites.
warthog (Flagstad et al., 2001; Muwanika et al., 2003) as well as carnivores, such as the African lion (Bertola et al., 2016) show signatures of deep genetic differentiation across this region, with distinct evolutionary trajectories for populations in sub-Saharan Africa north of the rainforest belt compared with those in Southern and East Africa. Many of these patterns have been moulded by repeated contractions and expansions of the forests during the Pleistocene, which have driven concomitant contractions and expansions in the ranges of species (Dupont, 2011). Water systems have similarly been affected by climatic cycles (Goudie, 2005). For instance, the Chad Basin, centred on Lake Chad, was repeatedly connected to the surrounding basins of the Niger, Nile and Congo rivers during wetter phases of the Pleistocene and blocked by dune formation during arid phases (Goudie, 2005; Shettima et al., 2018). The presence of the rainforest, when taken with the cyclic variations in forest and hydrographic system extensions, have certainly driven the oldest vicariant event in the roan, splitting our north-west lineage from the remaining ones around 630 kya. We postulate that the White Nile marks the current separation between the north-west and the north-east population groups, and that rearrangements in this water system that led to the connection of White and Blue Nile rivers across central Sudan during the late-Pleistocene (Goudie, 2005) may have contributed to the separation between these two groups.

Another well-known barrier driving differentiation in savanna-adapted species is the EARS, which remained active throughout the Pleistocene (Chorowicz, 2005) and drove habitat instability in the region. EARS has been seen as a major barrier to gene flow not only in large mammals, including ungulates such as waterbuck (Lorenzen et al., 2006), giraffe (Lorenzen et al., 2012; Winter et al., 2018) or sable antelope (Jansen van Vuuren et al., 2010; Vaz Pinto, 2018) but also for carnivores including African lion and cheetah (Bertola et al., 2016; Charruau et al., 2011). Well-characterized geomorphological instability in this region, such as the formation of Lake Victoria and the drainage reversal of the associated Lake Kyoga around 400 kya (Goudie, 2005) can explain the vicariant event ca. 340 kya separating our north-east and east-central mitogenomic lineages and leading to the establishment of the eastern and central population groups. Lastly, in Southern Africa, the Zambezi River has been described as a major driver of differentiation within mammals (e.g. McDonough et al., 2015; Vaz Pinto, 2018). Extreme hydrographic changes that took place during the middle Pleistocene in this region (Walford et al., 2005) led to the connection of the Upper and Middle Zambezi, likely contributing to the more...
recent vicariant event separating east-central and southern lineages ca. 240 kya.

Despite being highly effective as drivers of vicariance, these barriers have remained semi-permeable as evidenced through the presence of contact (or hybrid) zones in many species. The contact zone between our north-east and eastern roan antelope groups in Uganda, north of Lake Victoria, overlaps with contact zones described for hartebeest, kob antelope, and African buffalo (Arctander et al., 1999; Lorenzen et al., 2007; Smitz et al., 2013). Likewise, the contact zone between the eastern and central population groups (between Tanzania and Zambia) is mirrored in sable antelope and eland (Lorenzen et al., 2010; Vaz Pinto, 2018). Additionally, the sharp contact zone seen between roan antelope mitochondrial lineages in Cameroon is geographically close to a region where forest and savanna elephants interbreed (Mondol et al., 2015), although these authors argue that the hybrid zone in the elephant is recent and probably the result of habitat modifications and poaching. In roan antelopes, this contact zone could either be a result of a north-east mitochondrial introgression into the north-west group, or alternatively a predominant dispersal of north-west males towards the north-east group. Both phenomena could have been facilitated by expansions of savanna habitats during interpluvials.

4.2 Genetic structure and congruence with subspecies

Based on geographical range, Ansell (1972) recognized six subspecies for the roan antelope. The validity and geographical correspondence of these subspecies have been questioned by genetic data (Alpers et al., 2004; Matthee & Robinson, 1999), and are not currently recognized by the IUCN (East & IUCN SSC Antelope Specialist Group, 1999; IUCN SSC Antelope Specialist Group, 2017). In fact, no clear and consistent phenotypic differences exist across the range of the roan antelope (based on our own personal observations). Rather, the boundaries between subspecies appear to be largely defined by gaps in the ranges (such as between northern Uganda and Sudan, the boundary between H. e. bakeri and H. e. langheldi), geopolitical borders (such as between the Albertine Rift and African lake system that separates H. e. cottoni from H. e. langheldi). Disentangling this question requires extensive sampling covering areas representing the six subspecies, which is offered for the first time by our work. Notably, we found near-perfect congruence between our distinct genetic groups and five subspecies. For both nuclear and mitochondrial data, we did not find molecular evidence for the sixth subspecies, H. e. charicus, originally described from the area between Benin and Sudan. Rather, individuals from this region belong to the nuclear north-west group (corresponding to H. e. koba), while their mitogenomes cluster either in the north-west (H. e. koba) or north-east (H. e. bakeri) lineages. This finding is consistent across both our contemporary and historical samples. Hence, from our molecular results, we argue that the historically proposed six subspecies rather correspond to five evolutionary lineages in roan antelope.

4.3 Refugial areas, biotic drivers and management

Information derived from our summary statistics for the different roan antelope groups strongly hint at the placement of refugia as well as the potential evolutionary origin of the species. For the roan antelope, diversity statistics show a gradual decrease from the north, and particularly the north-west, towards the east and southern groups, supporting either a single Pleistocene refugium in West Africa (previously proposed by Alpers et al., 2004), or alternatively, two Pleistocene refugia both in the northern range, in north-west and north-east; we also speculate that West African roan antelopes are, or belong to, the most ancestral population based on divergence dates between groups. Our findings follow the classical scenario where serial founder effects during range expansions reduce genetic diversity and change allele frequencies, resulting in population structure (Excoffier & Ray, 2008; Excoffier et al., 2009). A West African origin for taxonomic groups has similarly been proposed for other ungulates including hartebeest, buffalo and giraffe (Brown et al., 2007; Flagstad et al., 2001; Smitz et al., 2013).

Notwithstanding, the roan antelope persisted in at least five refugia (corresponding to our nuclear groups) scattered across its range, maintaining constant effective population sizes. More recently, roan antelopes started to experience a general trend of population decline possibly triggered by severe Holocene droughts together with increased human-pressure at the beginning of the Neolithic, as happened for other water-dependent ungulate species such as the African buffalo (Heller et al., 2008, 2012).
More recently, despite the relatively specific habitat requirements (mesic savanna woodland and constant availability of standing water), sensitivity to inter-species competition, and shy nature, roan antelope numbers appear relatively stable across large areas of its range (East & IUCN SSC Antelope Specialist Group, 1999), whereas in many regions the species is virtually extinct, notably across its southern distribution. Roan antelope populations have been decreasing for the past three decades, due to a combination of factors largely ascribed to human impact (IUCN SSC Antelope Specialist Group, 2017), and the majority of populations are now found within protected areas. This situation creates an entirely new set of considerations, including the persistence of population diversity over time in small and/or closed populations. Although translocations add value as a management tool, especially when the individuals that are being translocated have high levels of heterozygosity (Scott et al., 2020) these actions should be undertaken only after a number of factors have been considered (see Jansen van Vuuren et al., 2017). Expanded surveys across the southern region are needed to fully access these concerns.

With increasing number of African antelope species facing anthropogenic threats (East & IUCN SSC Antelope Specialist Group, 1999), it is important to take their evolutionary histories into account when applying conservation management plans in the wild (Collen et al., 2011; Hartmann & André, 2013). African antelopes have been imprinted by the same climatic and geological pressures, sharing many refugial areas that dictated subsequent range expansion rates, and currently observed spatial patterns (Lorenzen et al., 2012). Shared evolutionary histories can determine common responses to current threats and, therefore, the continued existence of a large number of species.

ACKNOWLEDGEMENTS
Historical samples for this study were generously provided by the following five museums: Bulawayo Natural History Museum (Zimbabwe), Powell-Cotton Museum (UK), Berlin Natural History Museum (Germany), Wien Natural History Museum (Austria) and Swedish Natural History Museum (Sweden). We thank Pedro Vaz Pinto for donating contemporary samples from Angola. No field sampling was done on purpose for this study. We thank lab assistance from CIBIO technical staff, especially to Diana Castro and Susana Lopes. We acknowledge the Chief Editor Professor Rosemary Gillespie, the Editor Dr. Rayna Bell, and two anonymous reviewers for providing detailed suggestions that considerably improved the quality of our work. MG and RG were supported by the Portuguese Ministry of Education and Science, through the national project CISTER (National inBIO) and of the Ecological Genomics and Wildlife Conservation Research Group at the University of Johannesburg and co-funded by the project NORTE-01-0246-FEDER-000063, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Microsatellite genotype data used for the analyses in this manuscript were uploaded to the Open Science Framework repository under the https://doi.org/10.17605/OSF.IO/9E74F. Whole mtDNA sequences and the complete alignment generated in this work are available in GenBank (BioProject PRJNA682311; accession numbers MZ560849-MZ560930).

ORCID
Margarida Gonçalves https://orcid.org/0000-0001-8720-3959
Hans R. Siegismund https://orcid.org/0000-0001-5757-3131
Bettine Jansen van Vuuren https://orcid.org/0000-0002-5334-5358
Raquel Godinho https://orcid.org/0000-0002-4260-4799

REFERENCES


**BIOSKETCH**

**Margarida Gonçalves** is broadly interested in evolutionary biology of African species. This work represents a component of her PhD work on the evolutionary history of the *Hippotragus* genus. Using genetic and genomic tools, the goal is to understand historical population dynamics that shaped current population structure of both roan and sable antelopes.

**Author Contributions:** NF and RG conceived the ideas; MG, HRS and BJV collected field samples with additional material from collaborators; MG performed lab work and analysed the data with assistance from HRS, BJV and RG. MG led the writing with contributions from all co-authors.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.