The evolutionary history of the *Hippotragus* genus: integration within African phylogeography

Margarida Ladeira Felício Gonçalves

Doctoral Programme in Biodiversity, Genetics, and Evolution

Department of Biology

2021

Supervisor

Raquel Godinho, Principal Investigator CIBIO/InBIO,
Invited Assistant Professor at University of Porto

Co-supervisors

Hans Redlef Siegismund, Professor at University of Copenhagen

Bettine Jansen van Vuuren, Professor at University of Johannesburg

Margarida Gonçalves acknowledges the PhD scholarship PD/BD/114032/2015 granted by FCT - Fundação Portuguesa para a Ciência e Tecnologia.
This work was supported by National Funds through FCT in the scope of the project UIDB/50027/2020, as well as private funds of CIBIO/InBIO and of the Ecological Genomics and Wildlife Conservation Research group at the University of Johannesburg.

This thesis contains research studies, presented in the form of chapters, which correspond to scientific manuscripts that are either published, submitted or in preparation for future publication. Considering all studies were accomplished with the collaboration of several authors, namely the co-supervisors, Hans Redlef Siegismund (University of Copenhagen) and Bettine Jansen van Vuuren (University of Johannesburg), the student states that she actively participated in their conception, through sampling acquisition, data analysis, discussion of results, writing, and illustrations editing.

UNIÃO EUROPEIA
Fundo Social Europeu

Acknowledgements / Agradecimentos

In this long journey I did not walked alone.

I would first like to thank my supervisors, for all the support and assistance throughout this journey. For pushing me forward, into sharpening my scientific thinking and bringing my work to a higher level.
To Raquel, for the formulation of ideas, insightful feedback on the research questions, methodology, and conclusions. For her sharp response when dealing with setbacks, and humanity. For giving me the opportunity to be part of this amazing project and providing me the means to accomplish its goals.
To Hans, for the scientific input and insightful feedback onto my work, for his sharp opinions and forward thinking. For making me feel welcome when in Copenhagen and introducing me to his lovely family.
To Bettine, for the scientific input and insightful feedback onto my work, for her expert ideas and clarity. For making me feel welcome and safe when in Johannesburg and giving me the opportunity to experience the wonders of the African savanna for the first time.

To all collaborators that provided or facilitate the acquisition of valuable samples to this work, namely Dr. Inbal Livne (Powell-Cotton Museum), Dr. Frank Zachos (Natural History Museum Wien), Steffen Bock (Natural History Museum Berlin), Tsitsi Maponga (Natural History Museum Bulawayo), Dr. Daniela Kalthoff (Swedish Natural History Museum), as well as Dr. Teresa Lobo Fernandes and Dr. Rui Bernardino (Lisbon Zoological Garden).

To the members of my research group, for helpful input and support, namely Rita Rocha, Pedro Silva, Maris Hindrikson, Diana Lobo, Carolina Pacheco, Joana Rocha, Mariana Ribeiro, Margarida Andrade, and Gonçalo Ferraz.
To Prof. Nuno Ferrand for giving me the opportunity to be part of this amazing project.

To Klaus-Peter Kloepfli for contributing to this work. To Miguel Carneiro, Pedro Tarroso, Pedro Vaz Pinto, Stephen Sabatino, and Rasmus Heller for scientific input.
To the CTM technical staff for lab assistance, support, and friendship, namely Susana Lopes, Diana Castro, Patrícia Ribeiro, Sofia Mourão, Sandra Afonso, and Maria Magalhães.

Quero agradecer a todos os meus colegas e amigos que me apoiaram.

À Diana, pela sua amizade, força e apoio. Pelas conversas infindáveis, pelos desabafos, e por ter acalmado as minhas dúvidas, desesperos e desilusões.

Estou grata por a minha família.
Por a minha mãe, por tudo o que me deu. Por me ter mostrado que o amor não conhece tempo, espaço ou forma.
Por o meu pai, por tudo o que sou. Pelo amor, educação e valores que me transmitiu. Por nunca duvidar do que sou capaz e me fazer querer alcançar sempre novos sonhos.
Por a minha irmã Mariana, os meus sobrinhos e os meus tios, pelo amor e sorrisos.

E por fim, agradeço ao meu companheiro desta e de todas as aventuras, Pedro. Pelo amor, compreensão, paciência e força. Por ter partilhado esta loucura comigo e nunca ter deixado de acreditar.

Esta tese é dedicada ao meu tio-Avô Mário,
por todas as dicas que me deu sobre a Vida.

Resumo

Para melhor compreender a história evolutiva de um táxon, é importante avaliar a sua estrutura e diversidade genética espacial, considerando a história climática e geomorfológica do bioma que este habita. O bioma de savana Africana alberga a maior diversidade de ungulados na Terra, incluindo espécies de antílopes, cujo surgimento coincide com uma mudança climática a nível continental, durante o final do Plioceno, que resultou numa alteração de cobertura vegetal, com dominância de savanas. Os ciclos climáticos subsequentes do Pleistoceno foram marcados por repetidas contrações de savana durante períodos pluviais, húmidos e quentes, e expansões durante períodos inter-pluviais, frios e secos. Para taxa adaptado a este bioma, refúgios de zonas de savana permitiriam a persistência de populações durante os períodos pluviais, com expansões populacionais subsequentes durante períodos mais favoráveis, associados a inter-pluviais. A presença de barreiras geomorfológicas ditou a posição de zonas de contato entre populações, moldando os padrões espaciais atualmente observados. Esta tese é focada em duas das espécies mais emblemáticas de antílopes africanos, a palanca-vermelha (*Hippotragus equinus*) e a palanca-negra (*H. niger*), os únicos representantes vivos do género. Diversos estudos têm avaliado níveis de estrutura, diversidade e diferenciação populacional para estas espécies. Contudo, para a palanca-vermelha, estes são bastante limitados em termos de amostragem geográfica e resolução genética. Neste contexto, esta tese contém a avaliação filogeográfica mais abrangente até à data, para esta espécie. Usando marcadores genéticos e genómicos, são avaliados, ao longo da distribuição geral da palanca-vermelha, padrões genéticos espaciais e várias estatísticas ao nível intraespecífico, bem como são dadas estimativas de tempos de divergência e de tendências demográficas. Os resultados dão um melhor contexto genético para as subespécies reconhecidas e relacionam os padrões estruturais encontrados com as flutuações climáticas do Pleistoceno e barreiras geográficas atualmente reconhecidas em África. Ao completar o primeiro genoma de referência para esta espécie, esta tese também fornece recursos genómicos para pesquisas futuras e gestão. As palancas, como espécies adaptadas à savana, partilham muitas características biológicas e ecológicas, tornando-as excelentes modelos para estudos de genómica comparativa. Seguindo esta abordagem, os diferentes padrões espaciais encontrados para cada espécie podem ser relacionados com histórias evolutivas únicas, quando submetidas aos mesmos fatores extrínsecos, mas também a fatores intrínsecos distintos, nomeadamente especialização ecológica e sazonalidade reprodutiva. No sentido de contextualizar o clima e a história geomorfológica do continente Africano com a evolução do género *Hippotragus*, de outras espécies de

antílopes, bem como de restantes taxa adaptados à savana, esta tese apresenta também a primeira revisão sobre descontinuidades genéticas para espécies de fauna e flora associadas à savana Africana. Padrões concordantes a vários grupos taxonómicos sugerem trajetórias evolutivas comuns e permitem melhor identificar os mecanismos que moldaram a biodiversidade atual deste bioma tão importante.

Palavras-chave

África subsaariana, antílopes, ciclos climáticos do Pleistoceno, demografia, diversidade genética, estrutura populacional, genómica comparativa, filogeografia, história evolutiva, microssatélites, mitogenoma, montagem *de novo*, savanas, ungulados

Summary

Towards understanding the evolutionary history of a particular taxon, it is important to assess its spatial genetic structure and diversity against the climatic and geomorphological history of the biome it inhabits. The African savanna biome harbours the greatest diversity of ungulates on Earth, including antelope species, whose emergence coincides with a continental climatic shift during late Pliocene, that led to increased dominance of grasslands. During subsequent Pleistocene climatic cycles, this area was marked by repeated savanna contractions during warm and wet pluvials, and expansions during cold and dry interpluvials. For savanna-adapted taxa, grassland refugia would enable populations to persist during pluvials, with subsequent range expansions during favourable interpluvials. The presence of geomorphological barriers dictated the position of contact zones, shaping currently observed spatial patterns. This thesis is focused on two of the most emblematic African antelope species, the roan (*Hippotragus equinus*) and the sable (*H. niger*), the only extant representatives of the genus. Several studies have been assessing their levels of structure, diversity, and differentiation at the intraspecific level. For the roan antelope, however, such studies suffered from incomplete geographic sampling and poor genetic resolution. In this context, we provide the most comprehensive phylogeographic assessment for this species. Using both genetic and genomic makers, we assess the spatial genetic patterns across roan antelope' entire distributional range, and analyse several statistics at the intraspecific level, estimate divergence time intervals, and perform demographic analysis. Results provide a better genetic context for the recognized subspecies and relate structural patterns with the Pleistocene climatic fluctuations and recognized geographic barriers across sub-Saharan Africa. By assembling the first reference genome for this species, we also provide genomic resources for future research and management. As savanna-adapted species, the roan and sable antelopes share many biological and ecological characteristics, being powerful models for comparative genomics. Following this approach, resulting differential spatial patterns can be related to unique evolutionary paths in response to the same extrinsic factors, as well as distinct intrinsic factors, namely ecological specialization, and reproductive seasonality. Towards contextualizing African' climate and geomorphological history with the evolution of, not only the *Hippotragus* genus, but also other antelope species, and remaining savanna-adapted taxa, this thesis also presents the first review on genetic discontinuities for fauna and flora across the savanna biome. Similar patterns across several taxonomic groups suggest common evolutionary trajectories and a better identification of the mechanisms that have been shaping the biodiversity of this important biome.

Keywords

antelopes, comparative genomics, demography, *de novo* assembly, evolutionary history, genetic diversity, microsatellites, mitogenome, Pleistocene climatic cycles, population structure, phylogeography, savannas, ungulates, sub-Saharan Africa

Contents

List of Tables

Chapter 2

Chapter 3

Study I

Study II

Chapter 4

Chapter 5

List of Figures

Chapter 1

Chapter 2

Chapter 3

Study I

Study II

Chapter 4

Chapter 5

List of Abbreviations

a.s.l. – at surface level

ASG – Antelope Specialist Group

bp – base-pairs

ca. – about (*circa*)

CVL – Cameroon Vulcanic Line

DNA – Deoxyribonucleic acid

DRC – Democratic Republic of Congo

EAM – Eastern Arc Mountains

EARS – East African Rift System

e.g. – for example (*exempli gratia*)

ESUs – Evolutionary Significant Units

Gbp – giga base-pais

i.e. – that is (*id est*)

IUCN – International Union for Conservation of Nature and Natural Resources

Kbp – thousand base-pairs

kya – thousand years ago

mtDNA – mitochondrial deoxyribonucleic acid

Mya – million years ago

NCBI – National Center for Biotechnology Information

NGS – Next Generation Sequencing

PCR – polymerase chain reactions

RNA – Ribonucleic cid

rRNA – ribossomal Ribonucleic acid

SNPs – single nucleotide polymorphisms

spp. – species plural

ssp. – subspecies

SSC – Species Survival Commission

UK – United Kingdom

USA – United States of America

CHAPTER 1

General Introduction

1.1 Sub-Saharan Africa

Sub-Saharan Africa is the geographic area of the African continent that lies south of the Sahara Desert, comprising of different climatic zones and a wide variety of vegetational and bare-soil covers. This thesis is focussed on a particular biome of Africa, the savanna biome, which harbours the greatest diversity of ungulates on Earth (Du Toit and Cumming 1999), and in particular two antelope species that occupy this biome namely the roan and the sable antelope (*Hippotragus* spp.). To understand the evolutionary history of these savanna dwellers, attention must be drawn to extrinsic factors that have been shaping the savanna biome such as climate, vegetation cover, and geomorphological features.

1.1.1 Climate and land-cover

According to the widely used Köppen-Geiger climate classification system (Peel, Finlayson and Mcmahon 2007), sub-Saharan Africa is composed of three main climate types: tropical (group A), arid (group B), and temperate (group C) (Figure 1.1).

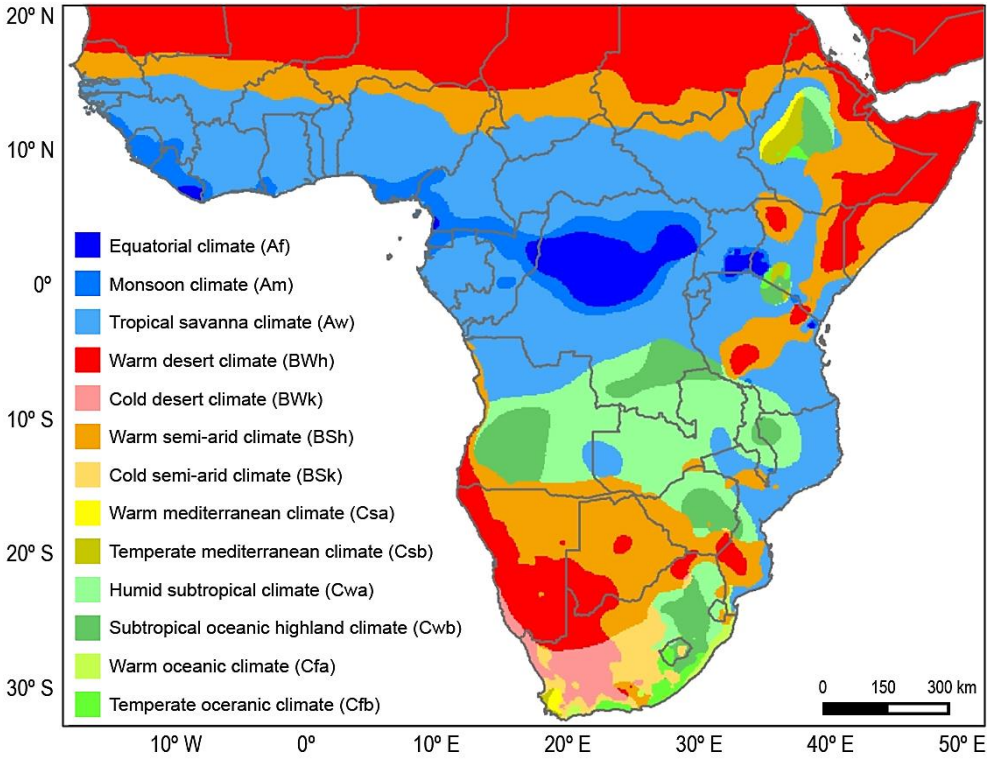


Figure 1. 1 Climate zones of sub-Saharan Africa, according to the Köppen-Geiger climate classification system, adapted from Peel et al. (2007), following the authors' colour-codes. Colours show the transition between tropical climates (blue), arid climates (red and orange), and temperate climates (yellow and green), as described in the text.

Tropical climate spans a wide portion of sub-Saharan Africa. It is mainly composed of tropical savanna climate, characterized by warm temperatures year-round and distinct wet and dry seasons. In Central Africa, across the Democratic Republic of Congo (DRC), there is a region of equatorial and monsoon tropical climate, composing the core area of the Tropical Forest belt. Arid zones in sub-Saharan Africa are represented by four main climate zones namely two warm and two cold. The warm desert climate is present in the Horn of Africa (across Ethiopia, Somalia, and north-eastern Kenya) and in the Kalahari Desert (across Namibia, Botswana, and north of South-Africa) and the warm semi-arid climate is located in the regions that surround the former two, as well as in the Sahel (the transitional zone between the Sahara Desert and tropical savanna). These climates tend to have hot to extremely hot summers and warm winters with some precipitation. The cold arid zones are mostly present in South-Africa, especially in the Northern Cape Province, and are characterized by cold desert and semi-desert climates, with hot dry summers and cold winters. Temperate zones exist mainly across Southern Africa, from central Angola through Malawi. These are characterized by humid subtropical and subtropical oceanic climates, with warm humid summers and mildly cold winters. Other pockets of this type of climate occur throughout Afromontane forests, across the Ethiopian Highlands to north-eastern South-Africa. In the Cape region of South-Africa there is also a pocket of warm Mediterranean climate, characterized by moderate year-round temperatures and rainy winters (Peel et al. 2007).

The distribution of natural vegetation is mainly determined by climate (most notably temperature and rainfall) and it plays a key role in biological processes, acting as source and sink in biogeochemical cycles (Sellers et al. 1997). Four main groups of natural land-cover types make up sub-Saharan Africa: forests; woodlands and shrublands; grasslands; and bare soil, including desert sands, rocky areas, and waterbodies. However, anthropogenic actions can lead to permanent changes on the land-cover, by turning natural vegetation into croplands, and a fifth group of land-cover, including non-natural agricultural land, was included in the updated land-cover map of Africa by Mayaux et al. (2004) (Figure 1.2), which takes into consideration the work of White (1983) on assessing vegetation types across the continent.

The Tropical Forest belt, across Central Africa, is mainly composed of evergreen closed forests, with more than 70% tree cover at lower altitudes (below 900 m at surface level – a.s.l.). It thrives across southern Cameroon, Equatorial Guinea, Gabon, Congo, and most of DRC, under tropical climate. Montane evergreen forests occur at higher altitudes (above 900 m a.s.l.), mostly located along the Rift Valley, in East Africa. Deciduous and mixed forests of equal tree to shrub cover, make the transition from closed forests to woodlands and shrublands, through the continent. Regions with less

than 40% tree cover and an increased percentage of shrubs and grasses are dominated by woodlands, shrublands and grasslands. Woodlands and shrublands occur mostly throughout tropical savanna climate, whereas grasslands are associated with reduced precipitation (warm and cold semi-arid and desert climates) and tree cover falling below 20% or even becoming absent. When these grasslands become regularly flooded, they make up swamplands, an example of which is located across Sudan. Non-vegetational land includes sandy desert and dunes, stony deserts, and bare rock areas, mostly found along the coastline of Namibia. Remaining areas consist of waterbodies, the main ones composing the Great Lakes, in East Africa. Human-altered landscapes encompassing agricultural land and cities, comprise about 11% of the total land in Africa, especially throughout West and East Africa (Mayaux et al. 2004).

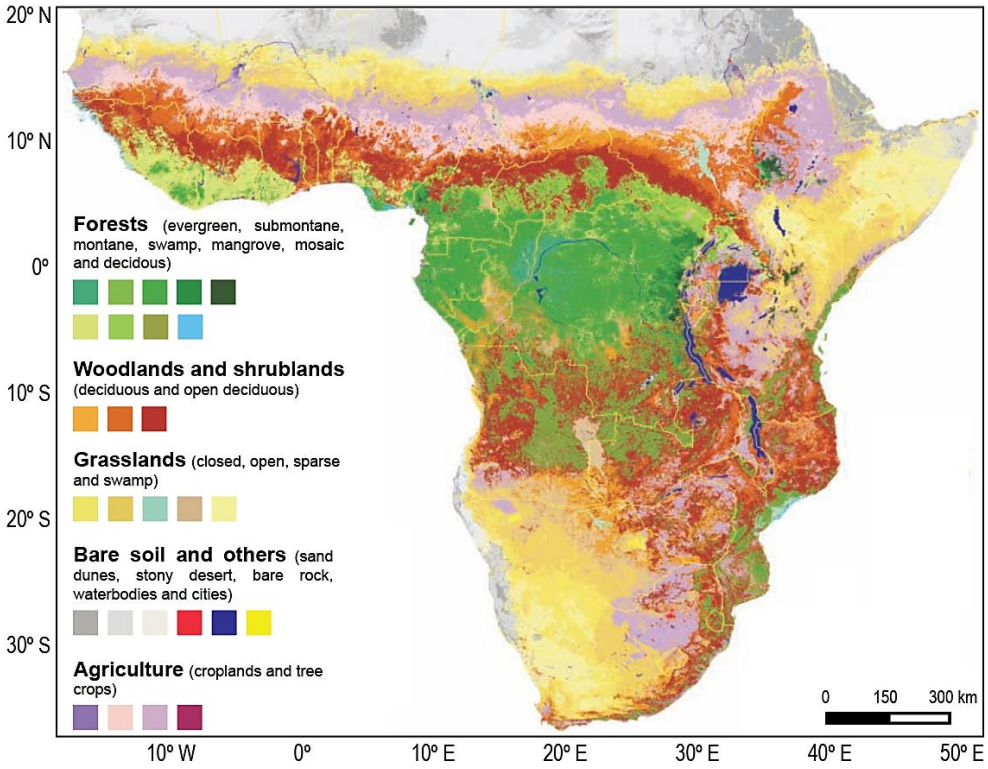


Figure 1. 2 Land-cover map of sub-Saharan Africa, adapted from Mayaux et al. (2004), following the author's colour-codes. Main natural covers depicted are: forests, woodlands and shrublands, grasslands, and bare soil. A fifth cover consists of agricultural land.

In particular, the African savanna biome is therefore characterised by a mixed system of woodlands, shrublands and grasslands, with open canopy. It occurs mostly throughout tropical savanna and humid tropical climates, with warm temperatures year-round and seasonal availability of water (White 1983). Some African antelopes are dependent on this type of biome, feeding mainly on savanna grasses. On the other hand, woodlands, shrublands and grasslands are dependent on mammalian herbivory to maintain an open

canopy (Asner et al. 2009), creating a positive feedback. Other than herbivory, this positive feedback is also maintained by fire. A higher density of grasses enables frequent fires, which, in turn, prevents the outgrowth of tree canopy (Keeley and Rundel 2005; Beckage, Platt and Gross 2009). The balance of the savanna biome is therefore delicate and human activity has altered vegetation types, by converting large areas into croplands and tree crops (Asner et al. 2004). At the landscape scale, this creates a mosaic of non-native vegetation along discontinuous patches of suitable habitat for savanna-adapted fauna, such as some antelope species.

1.1.2 Geomorphological features

Topographically complex regions are typically associated with high species richness and spatial turnover, by creating new habitats, expanding environmental gradients, and establishing barriers for animal dispersal (Badgley 2010). The African continent has a unique bimodal topography, with the south-eastern parts of sub-Sahara comprising a plateau 1,000 m a.s.l., contrasting with the northwest, much of which remains at 500 m a.s.l. This contrasting physiography resulted from the break-up of the Gondwana during the Mesozoic, 120 to 180 million years ago (Mya) with continued uplift of the continent until the end of the Cretaceous, about 65 Mya (Doucouré and de Wit 2003). Drastic landmass configurational changes persisted throughout the early Miocene (~30 Mya) until today, shaping African landscape through processes of continental uplift and tilting, remarkable drainage evolution, as well as continuous rifting and volcanism across East and Southern Africa (Livingston and Kingdon 2013).

Continent moving is driven by deep mantle currents, which manifest on the surface as domes (King and Ritsema 2000). When domes split open, it leads to tectonic basins or rift valleys, which become gradually wider by divergent movement of continental plates. Several domes exist today in Africa, with the widest and most extensive one located in East Africa (Pik, Marty and Hilton 2006). The East African Rift System (EARS) comprises one of the most spectacular geological features across the continent. In a series of successive aligned individual tectonic basins (or rift valleys), it extends for several thousands of kilometres through East and Southern Africa. Each basin is controlled by several tectonic faults, extending up to tens of kilometres wide. The rift valleys are grouped into two main lines, forming the eastern and western (or Albertine) branches of the EARS (Figure 1.3). A third, smaller branch, extends to the southeast, along the western coast of Mozambique, consisting of an early stage, more recent rift. The eastern branch, with a distance of 2,200 km, runs from the Afar triangle, at the junction between the African, Arabian and Somalian tectonic plates, through the

Ethiopian and Kenyan (Gregory) Rifts, until central Tanzanian basins, consisting of the Eastern Arc Mountains (EAM). There is a sharp morphological contrast between the Ethiopian and Somalian plateaus, on either side of the Ethiopian Rift. Both plateaus rise up to 3,000 m a.s.l., while the highest mountains can reach 4,000 m a.s.l. across the Ethiopian highlands. Higher elevations on the Ethiopian side are, however, related to another uplift system, forming the Tana dome (Chorowicz 2005).

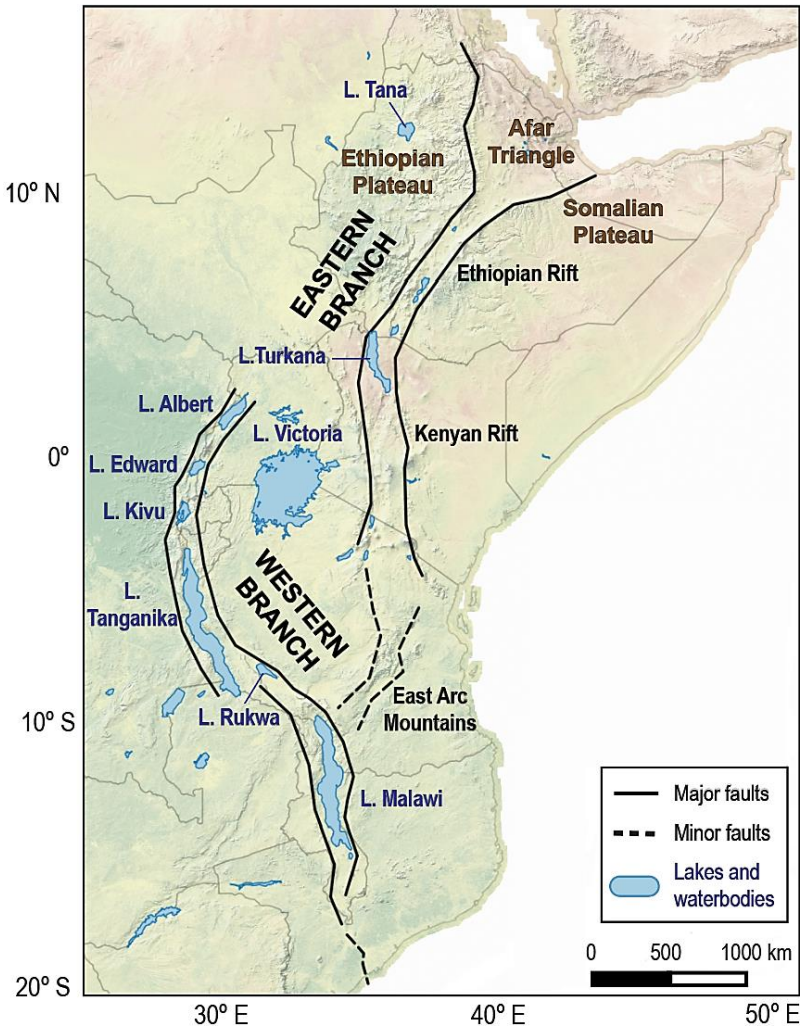


Figure 1. 3 Geographic representation of the East African Rift System – EARS, adapted from Chorowicz (2005). Major faults are depicted in black curves, representing the eastern and western branches, as well as minor faults, in black dotted curves, namely the Eastern Arc Mountains – EAM. Lakes and waterbodies are depicted, with corresponding names in blue, identified by "L.". Land cover is represented from forests (in green) to mountains (in brown), with shaded relief, depicting higher altitude.

The western branch, also called the Albertine Rift, extends for 2,100 km, from Lake Albert, located between the DRC and western Uganda, to Lake Malawi, across central Mozambique. Besides these, it also comprises Lake Edward, Lake Kivu, Lake Tanganyika, and Lake Rukwa, between Uganda and southern Tanzania (Chorowicz 2005). Most of the Great Lakes in Africa are, therefore, located along the EARS. The

only exception is Lake Victoria, located at a lower altitude valley between mountainous chains arising from the formation of both branches. Tectonically active volcanoes exist along the EARS, mainly across the wider Ethiopian dome, but also along the Kenyan and Tanzanian dome, to the south (Chorowicz 2005; Abebe et al. 2007), with the highest of them being the well-known Mount Kilimanjaro, with an altitude of almost 6,000 m a.s.l. Rift formation is dated back to the Mio-Pliocene periods, around 30 Mya, with the open fraction of the Ethiopian dome at the Afar triangle (Chorowicz 2005; Roberts et al. 2012). However, most of the modern topography of the EARS was generated throughout the last two million years, during the Pleistocene (Delvaux et al. 2012).

Domal uplifting is also responsible for most of the basin and swell topography of Africa, which subsequently had a profound effect on the continental drainage patterns (Moore and Blenkinsop 2002). Drainage systems in Africa have underwent distinct reorganizations since the early Cretaceous (~113 to 145 Mya) (Stankiewicz and de Wit 2006; Livingston and Kingdon 2013), resulting in eight main basins throughout sub-Saharan Africa today (Figure 1.4), with contemporaneous river courses being relatively recent (Goudie 2005).

Three main drainage systems are present in the north of the African tropical region. The Niger drainage system extends more than 4,000 km across West Africa, from the Guinea highlands, towards Mali, Niger, Benin, and Nigeria, where it ends in a massive delta, into the Atlantic Ocean. The Niger River has the Benue River as its main tributary, flowing across eastern Nigeria. The Nile drainage system includes the largest river in Africa, the Nile River, with nearly 7,000 km long, with two main tributaries, the White Nile, and the Blue Nile. The White Nile drains from Lake Victoria across Uganda, while the Blue Nile, the source of most of the Nile waterflow, begins at Lake Tana, in northern Ethiopia. Both tributaries merge across central Sudan, forming the Nile River that has its large delta in Egypt, into the Mediterranean Sea. The modern Nile River was only formed during the late Pleistocene, around 25 thousand years ago – kya. Before that, this area was drained by at least three different drainage systems, evolving through tectonic uplifts, and changing sea-level from the late Eocene (~40 Mya). Much of the current characteristics of the river are related to the development of the EARS, namely the uplift of the Ethiopian highlands and development of major lakes (Goudie 2005).

The African tropical region embraces the Congo drainage system, with the Congo River extending for more than 4,000 km across Central Africa and having the highest discharge flow in the continent. It flows from southern DRC, through the Tropical Forest belt, and major tributaries consist of the Great Lakes in East Africa, as well as other drainages from the highlands and mountains of the EARS. The centre-oriented flow of its main tributaries creates a leaf-like pattern that is unique in Africa, ending across the

Atlantic Ocean, without forming a delta. It is possible that before the Pleistocene, the Congo basin was occupied by large waterbodies, of which the lakes Tumba and Mai-Ndombe, are the sole remnants (Goudie 2005).

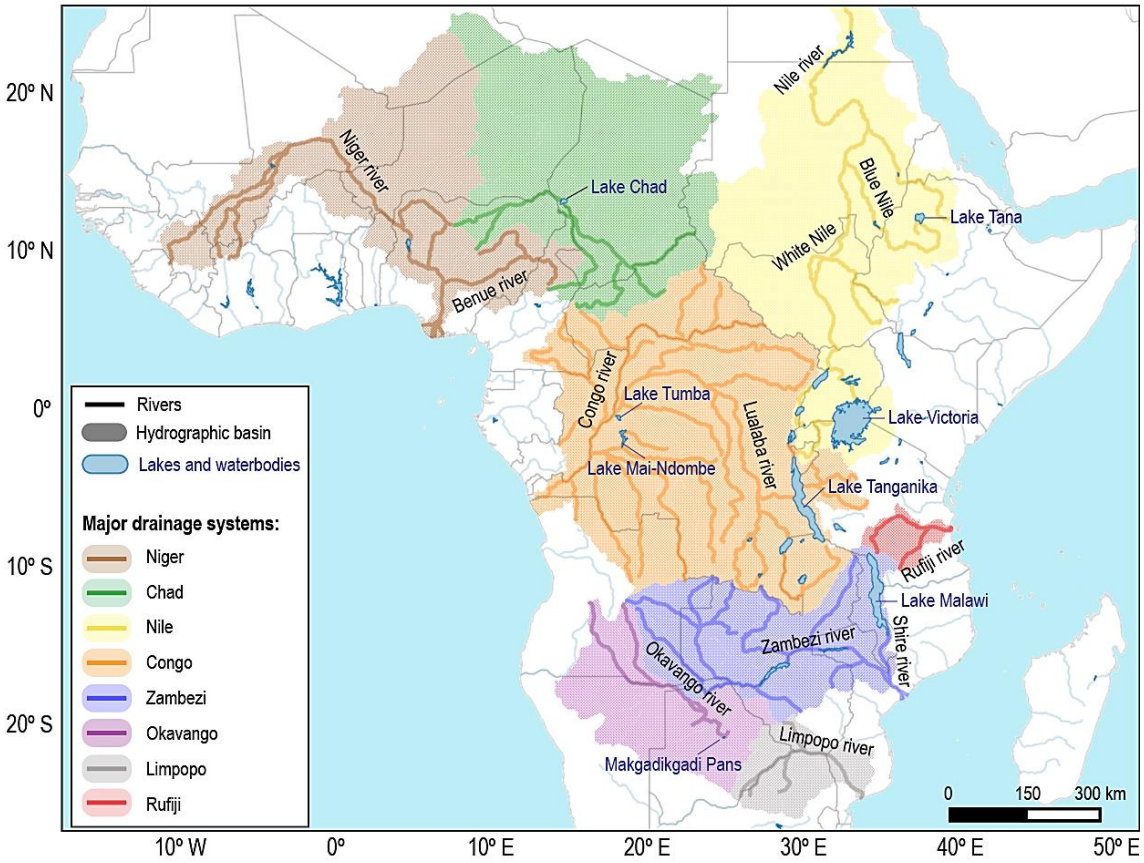


Figure 1. 4 Geographic distribution of major drainage systems across sub-Saharan Africa. Each hydrographic basin and corresponding rivers and main tributaries, are identified by a colour, as portrayed in the legend. Main rivers are identified (in black), as well as major lakes and waterbodies (in blue). Hydrographic basins retrieved from GeoNetwork (ww.fao.org).

South of the tropical region, there are three major drainage systems. The Zambezi River is about 2,500 km long, rising from northwest Zambia, through eastern Angola, northern Botswana, and northern Zimbabwe, into Mozambique, where its delta enters the Indian Ocean. It suffered extreme hydrographic changes through recent times, and only during Pliocene to mid-Pleistocene (~1 to 5 Mya) it became a joint system. Before that, the upper Zambezi was connected to the Limpopo drainage system, located further south, between the border of Zimbabwe and South-Africa, whereas the middle Zambezi was linked to the Shire River, across Malawi (Goudie 2005). The Okavango drainage system includes the Okavango River, which runs for about 1,600 km, across south-eastward Angola, through Namibia, and Botswana. It is an endorheic system that has its interior delta running into a swamp across the Kalahari Desert. During the rainy season, it flows into the extensive wetlands of the Makgadikgadi Pans, creating a seasonal

hotspot of biodiversity in the area (McCarthy and Ellery 1998). The development of the Okavango drainage is most likely associated with an extinct paleolake Makgadikgadi around the Kalahari, during late Pleistocene (~100 kya) (Moore, Cotterill and Eckardt 2012). The Okavango, Limpopo, and upper Zambezi rivers were thought to be connected via such paleolakes across the Kalahari (de Wit 1999; Moore and Larkin 2001; Goudie 2005; Burrough and Thomas 2008; Burrough, Thomas and Bailey 2009).

1.1.3 Pleistocene climate oscillations

Similar to previously described landmass configurational changes, the African continent also experienced cyclic climate oscillations throughout its history. The early Pleistocene (~1 to 2.8 Mya) witnessed the growth and subsequent expansion of icesheets in higher altitudes over the planet, and the onset of the Ice Age (Jansen and Sjøholm 1991; Raymo 1994). However, climate change at lower altitudes and their impact across tropical and sub-tropical habitats, remained poorly understood (Lister 2004). With the advent of deep-sea drilling, it was possible to recover marine sedimentary sequences, with well-dated records of eolian dust variability (Arimoto 2001; Prospero et al. 2002), as well as preserved pollen records. These palynological records, when compared with contemporaneous marine and terrestrial values, allow the inference of vegetation changes over several climate cycles (Dupont et al. 1999; Dupont and Wyputta 2003; Hooghiemstra et al. 2006). These paleoclimate records, together with statistical climate models, provided the means to understand how Pleistocene climate oscillations influenced African terrestrial habitat and the evolution of African taxa (DeMenocal and Rind 1993; Guilderson, Fairbanks and Rubenstone 1994; DeMenocal 1995).

The late Miocene (~7.5 to 11.5 Mya) and early Pliocene (~3.5 to 5 Mya) periods were characterized by moister climates that we see today in Africa, leading to expansion and diversification of tropical habitat and tropical adapted species (DeMenocal 1995; Moorley and Kingdon 2013). Through the late Pliocene (~1.5 to 3.5 Mya), however, higher trade wind activity led to a continental-wide aridification in Africa (DeMenocal 1995, 2004). Such climatic change resulted in a gradual change of vegetational cover, from tropical habitats into the extension of open woodlands and savannas (Dupont 2011). This period, characterized as a savanna-optimum, led to the expansion and major diversification of savanna-adapted taxa, including the antelopes associated with to this biome (DeMenocal 2004; Vrba 1996).

However, owing to changes in the Earth's orbital forcing parameters, and consequent variations in temporal and spatial distribution of solar radiation, African climates became

increasingly variable during the Pleistocene (DeMenocal 1995; Trauth et al. 2003, 2007; Trauth, Larrasoaña and Mudelsee 2009). This is well documented by the number of Marine Isotope Stages (MIS), which reflect more than 100 alternating climate cycles during the Pleistocene and Holocene (Wright 2000). In sub-Saharan Africa, these cycles were generally short, ranging from 10 thousand to 100 thousand years, alternating between cold and dry interpluvials, corresponding to glacial intervals, and warm and wet pluvials, corresponding to interglacial intervals. The latitudinal arrangement of forest and savanna biomes were conserved throughout climate cycles. However, forest vegetation increased during warm and wet pluvials, whereas during cold and dry interpluvials, they receded (Dupont 2011). A robust record of these forest oscillations come from pollen records of trees and grasses, with alternate agglomeration between pluvials and interpluvials, respectively. For example, grass pollen records of the Poacea family indicate northward shifts in the boundary between Sahara and the Sahel, as well as southward shifts between the savannas across Angola and Namibia, and the Namib desert during interpluvials. This indicates a larger extension of savanna biome across sub-Saharan Africa during interpluvials, stronger during the last two glacials (MIS 2-4 and MIS 6), through late Pleistocene into Holocene (Dupont 2011). Pleistocene alternating cycles would then influence the vegetation cover of Africa's major biomes, and transitional zones, such as semi-desert, would expand during transitional periods between a glacial and interglacial maximum (Dupont 2011). Each time changing climates shifted or altered the vegetation cover across a specific area, its inhabitants would become engulfed by a sub-optimal habitat, being forced to change their range, contract it, or adapt to a new set of conditions (DeMenocal 2004; Moorley and Kingdon 2013).

Changes across humidity, temperature, and solar radiation would also impact river flow intake and lake water levels across Africa (Trauth et al. 2003, 2005, 2007; Goudie 2005). Lacustrine sediment analysis attest to the impact of climate cycles across African drainage systems. During the arid phases of Pleistocene interpluvials, the expansion of the Sahara Desert southwards, as well as the Kalahari and Namib deserts northwards, led to the development of major dune fields, which would then guide the pattern of river courses through subsequent humid pluvials (Moore and Larkin 2001; Shaw and Goudie 2002; Goudie 2005; Dupont 2011; Lézine et al. 2011). The Niger River is thought to have flown north-eastwards, towards the Sahara, until it became blocked by these sand dune fields during interpluvials (Goudie 2005). On the contrary, during pluvial phases of both the Pleistocene and Holocene (~5 to 100 kya), the Niger drainage catchment would overflow further north, fed by a massif paleolake called Megachad, located where Lake Chad is today (Goudie 2005; Livingston and Kingdon 2013; Armitage, Bristow and Drake 2015). More to the northeast, the arid phases of the Pleistocene would turn seasonal the

link between the Nile River and the Great Lakes. This would have been mainly caused by the reduction of surface level of Lake Albert, interrupting the discharge of water into the White Nile, as well as the desiccation of Lake Tana, where Blue Nile begins today (Lamb et al. 2007; Livingston and Kingdon 2013). Due to major droughts occurring through late Pleistocene and early Holocene, some lakes even became closed saline systems. Lake Malawi's watersheds were converted into a semi-desert, Lake Tanganyika suffered a 95% water volume reduction, and Lake Victoria only reconnected to the Nile River at about 13 kya (Beuning et al. 2002; Cohen et al. 2007; Scholz et al. 2007; Stager and Johnson 2008; Lyons et al. 2015). During wetter phases, the Nile River flow would be augmented by currently extinct tributaries, via the Sobat River across Sudan, fed by paleolakes that formerly existed in the region (Pachur and Rottinger 1997; Maxwell and Haynes 2010). In Southern Africa, wetter phases probably reinforced the connection of Okavango, Limpopo, and upper Zambezi rivers, through expanding wetlands. These would subsequently feed the Makgadikgadi paleolake across the Kalahari (Burrough et al. 2009; Moore et al. 2012).

1.1.4 Biogeographical regions

The geographical portioning of Africa according to distinct biological groups was first proposed by White (1983), based on vegetation. It was recently updated by Linder et al. (2012), using multivariate statistics on taxa distribution, comprising vertebrate animals (mammals, birds, reptiles, and amphibians) and plants. Congruence across taxa can be associated with vegetation distribution, common responses to climate variation, as well as shared evolutionary history. Apart from the Saharan region, six biogeographical regions are coherently recognized across sub-Sahara: 1) Sudanian; 2) Congolian; 3) Ethiopian; 4) Somalian; 5) Zambezian; and 6) Southern Africa (Figure 1.5). These can be further divided into several sub-regions, depending on the taxa occurring in the region. Transitional regions exist between major biogeographical regions, which are characterised by high levels of species turnover, due to wider ecological tolerances. Tropical forests fall within the Congolian region, across Central Africa. The savanna biome is divided between the Sudanian and the Zambezian biogeographical regions, with distinct characteristics (see below). The Ethiopian biogeographical region, across Northeast Africa, acts as a transitional region between both savannas, being indistinguishable from the Sudanian region for mammals (Linder et al. 2012). Somalian is a low turnover species region, associated with arid climate across the Horn of Africa. Finally, Southern Africa biogeographical region falls on the southern bank of the Zambezi

River, and is associated with more temperate climate and higher specific diversity (Linder et al. 2012).

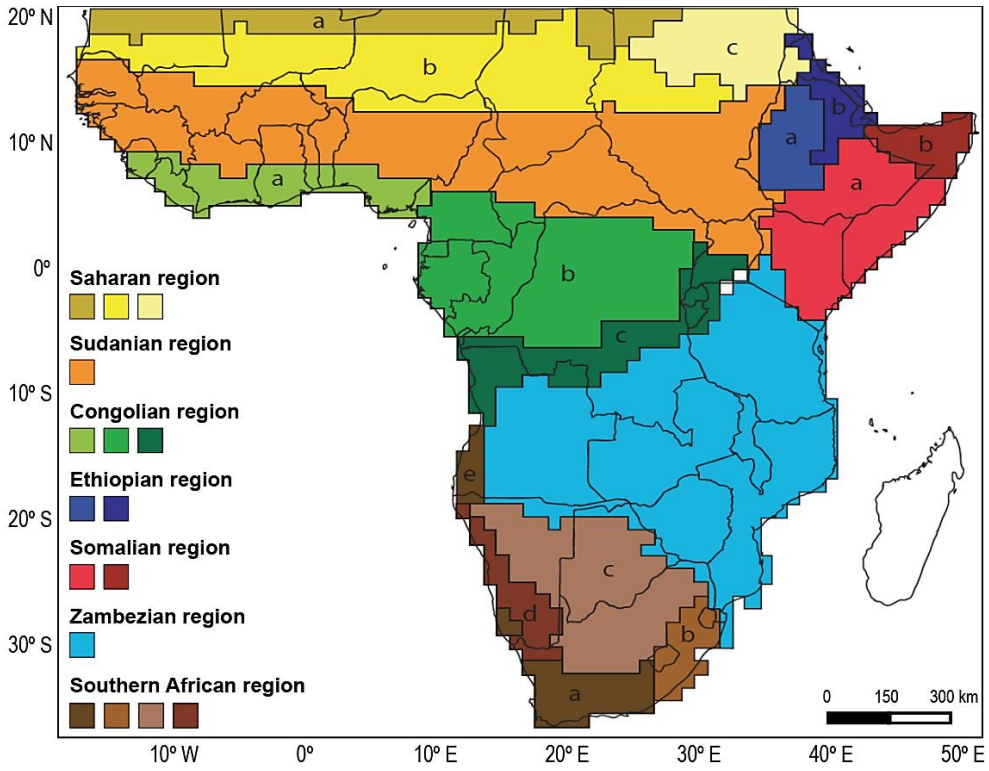


Figure 1. 5 Biogeographical regions across sub-Saharan Africa, adapted from Linder et al. (2012), based on the combined matrix of observed taxa (mammals, birds, reptiles, amphibians and plants). Seven major regions are recognized: Saharan, in yellow (subregions: a – Sahara; b – Sahel; c – Nubian Desert); Sudanian, in orange; Congolian, in green (subregions: a – Guinea; b – Congo; c – Shaba); Ethiopian, in dark blue (subregions: a – Djibouti; b – Ethiopia); Somalian, in red (subregions: a – Somalia; b – Horn); Zambezian, in light blue; and Southern Africa, in brown (subregions: a - Cape; b - Natal; c - Kalahari; d - Namib; e - south-west Angola). Regions designation and colours following Linder et al. (2012).

The Sudanian biogeographical region, one of the two regions comprising the savanna biome, extends from the coast of Senegal to the foothills of the Ethiopian Highlands, at the western part of the EARS. It is a low altitudinal region, with mean surface level below 750 m a.s.l., and main drainage system comprising the Nile-Chad rivers. Natural vegetation comprises wide extensions of woodland, mostly from genera *Isoberlinia* and *Acacia* (now referred to *Vachelia* and *Senegalia* genera, after a major taxonomic revision by Kyalangalilwa et al. 2013) at drier areas, and *Terminalia*, associated with seasonally inundated river valleys (White 1983). The savanna biome is also present in the Zambezian biogeographical region, which expands from southern Kenya towards Angola and Zimbabwe. It is mostly occupied by the Great African Plateau, with altitude above 900 m a.s.l. Wide water basins compose the Great Lakes, associated with the Great Rift Valley and main drainage system comprises the Zambezi River. The Zambezian and Sudanian regions share a quarter of their natural vegetation, however, the first has a

more diversified and richer flora, composed of evergreen forests at the northern limit with the Congolian region, woodlands and shrublands across the centre, and grasslands towards the southwest. Miombo woodlands are the most important vegetation type across the Zambezian region, associated to nutrient-poor soils and dominated by deciduous trees of the genera *Brachystegia*, *Julbernardia* and *Isoberlinia* (White 1983; Timberlake and Chidumayo 2011). East Africa presents a complex biogeographic pattern, in response to climate transition and topographic evolution of the EARS formation. The Rift, crossing Ethiopia, Kenya, Uganda, and Tanzania, is localized within four out of the six recognized regions for sub-Saharan Africa, associated with high species turnover across all observed taxa (Linder et al. 2012).

1.2 Antelope Evolution

Antelope is a term used to refer to several ungulate species within the Bovidae GRAY 1821 family, from the sub-order Ruminantia SCOPOLI 1777, order Artiodactyla OWEN 1848. They share many anatomical and behavioural characteristics, playing a central role in the nutrient cycling of the ecosystems they inhabit, where they shape the vegetation cover through herbivory and secondary seed dispersal, and are preyed by large carnivores (Hassanin et al. 2012; Bro-Jørgensen 2016). The classification of sub-families and tribes within the Bovidae is still debatable, with some alternatives proposed. Nevertheless, 88 antelope species exist today, classified into almost 30 different genera, most of which are native to Africa (Hassanin et al. 2012; Bro-Jørgensen 2016).

1.2.1 Shared characteristics

All antelopes share several behavioural and morphological adaptations that are taxonomically diagnostic (Brashares, Garland and Arcese 2000). They have hooves, meaning the tip of their toes is covered by a keratin cover that grows continuously (Prothero 1993; DeMiguel, Azanza and Morales 2014). Also, they have specialized and well-developed molar teeth, with a consequent reduction or even absence of canine and incisor teeth (Mendoza and Palmqvist 2008). Both sexes have bony cranial appendages (horns), which grow continuously. The horn size and shape vary greatly among species, but horns are always unbranched and have a bony core covered by a permanent sheet of keratin (Prothero, 1993; DeMiguel et al. 2014). All antelopes are even-toed ungulates, meaning that they bear there weight equally on the central two toes (Matthee et al. 2001). Also, as ruminants, all antelopes are able to digest pant-based matter in a four-chambered stomach (Matthee et al. 2001; Hume 2002). Most antelopes are mixed

feeders, browsing and grazing, according to seasonal availability of each plant type (Pérez-Barbería, Gordon and Illius 2001). Other shared morphological characteristics include a blunt-end snout, pointing ears, and long tail (Figure 1.6).

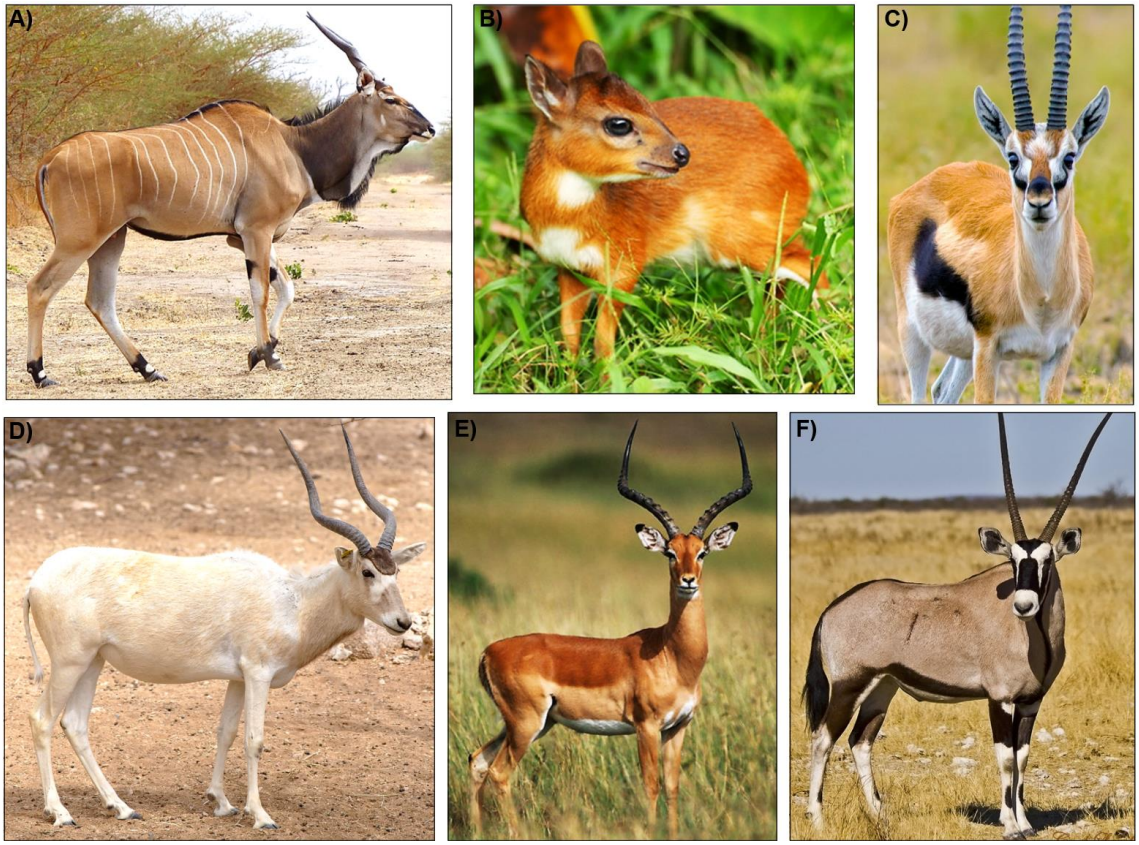


Figure 1. 6 Examples of African antelope species. A) giant eland (*Taurotragus derbianus*), photo by David Mallon (IUCN); B) royal antelope (*Neotragus pygmaeus*), copyright Harry Kouwer (britannica.com); C) Thompson's gazelle (*Eudorcas thomsonii*), photo by Yathin S. Krishnappa; D) addax (*Addax nasomaculatus*), photo by Jeff Kigma/Alamy (britannica.com); E) impala (*Aepyceros melampus*), photo by Frans Lanting (franslanting.photoshelter.com); and F) gemsbok (*Oryx gazelle*), copyright Carl Landsberg (treknature.com).

Antelopes vary greatly in size, both in height and weight. The smallest antelope species is the royal antelope (*Neotragus pygmaeus*), measuring 25 cm tall and weighing about 3 kg, whereas the largest one is the giant eland (*Taurotragus derbianus*), with a full mature male reaching 180 cm at the withers and weighing up to 1,000 kg. Most African antelopes have medium to large-sized bodies and present a clear sexual dimorphism, with larger and heavier males (Weckerly 1998; Loison et al. 1999; Pérez-Barbería, Gordon and Pagel 2002). Coat colouration ranges from pale white (as in the Addax – *Addax nasomaculatus*) to black (as in the sable antelope – *Hippotragus niger*), but intermediate brownish or reddish tones are more common. Across several species, sexual dichromatism leads to light-coloured coat across females and young males (Stoner, Caro and Graham 2003). Camouflage is also an important adaptation since all

species are prey for carnivores (Nokelainen and Stevens 2016). Among antelopes, it can manifest as prominent or faint stripes (as in the greater kudu – *Tragelaphus strepsiceros*), strong facial markings to hide the prominent eyes (as in gazelles – *Gazella* spp.), as well as blending body patterns (as in impala – *Aepyceros melampus*) or, on the contrary, strong contrasting colour patterns (as in the gemsbok – *Oryx gazella*), to conceal it from the background and confuse predators (Stoner et al. 2003). Most African antelopes are gregarious and territorial. Groups have a strong social structure, generally consisting of a territorial male and a variable number of females and juveniles. When males become sexually mature, they are forced to disperse into new territories, and fight other males for dominance, using their horns (Bro-Jørgensen 2011). Females probably developed their horns for defence against predators or express territoriality (Estes 1991; Stankowich and Caro 2009).

1.2.2 Phylogenetic context

Combining molecular, morphological, and fossil information obtained across several studies, Fernández and Vrba (2005) obtained a well-supported phylogenetic supertree for Ruminantia. Within Ruminantia, Bovidae constitutes a monophyletic family with at least nine sub-families. Incomplete fossil record, extensive morphological convergence, and a rapid radiation within Bovidae has led to taxonomic inconsistencies at the sub-familial level (Allard et al. 1992; Janecek et al. 1996; Hassanin and Douzery 1999; Matthee and Davis 2001). Nevertheless, two main clades have consistently been retrieved, comprising of the Bovinae and remaining taxa. Bibi (2013) and Hassanin et al. (2012) were able to retrieve concordant phylogenetic relationships analysing whole-mitochondrial data of extant Bovidae taxa. Recent genomic data corroborates previous reconstructions, while improving sub-familial inconsistencies (Chen et al. 2019) (Figure 1.7). Divergence time estimations across both the mitochondrial and nuclear genomes, together with fossil calibrations, agree with a general view of a bovid radiation during the Miocene through the Pliocene period. Estimated divergence time between Bovinae and non-Bovinae taxa is about 16 Mya, for both the nuclear and mitochondrial genome (95% CI 15.1-17.3 Mya; Bibi 2013; Chen et al. 2019).

The Bovinae clade includes the African buffalo (*Syncerus caffer*), and several African antelope species such as the greater kudu (*Tragelaphus strepsiceros*) and the giant eland (*Taurotragus derbianus*). The non-bovine clade also contains non-antelope African bovids, such as the wildebeest (*Connochaetes* spp.), as well as savanna-adapted antelopes, such as the Thompson's gazelle (*Eudorcas thomsonii*), the impala (*Aepyceros melampus*), and the hartebeest (*Alcelaphus buselaphus*). The Hippotraginae

sub-family comprises of antelope species only, from three different genera. Two of them – *Addax* and *Oryx* – include desert or semi-desert adapted-species. The first includes a single species, the addax (Addax nasomaculatus), distributed across the Sahara Desert. *Orix* genus contains four species: the East African oryx (*O. beisa*), across the Somalian region; the scimitar oryx (*O. dammah*), widespread throughout the Sahara; the gemsbok (*O. gazella*), distributed throughout arid regions of Southern Africa and the Kalahari Desert; and the Arabian oryx (*O. leucoryx*), restricted to the Arabian Peninsula. The third genus, *Hippotragus*, contains two extant savanna-adapted species: the roan (*H. equinus*) and the sable antelope (*H. niger*). This genus will constitute the main focus of this thesis. The node corresponding to Hippotraginae sub-family is estimated between 6.6 Mya, according to the mitogenome (95% CI 5.6-7.6 Mya; Bibi 2013), and around 8 Mya, according to the nuclear genome (Chen et al. 2019).

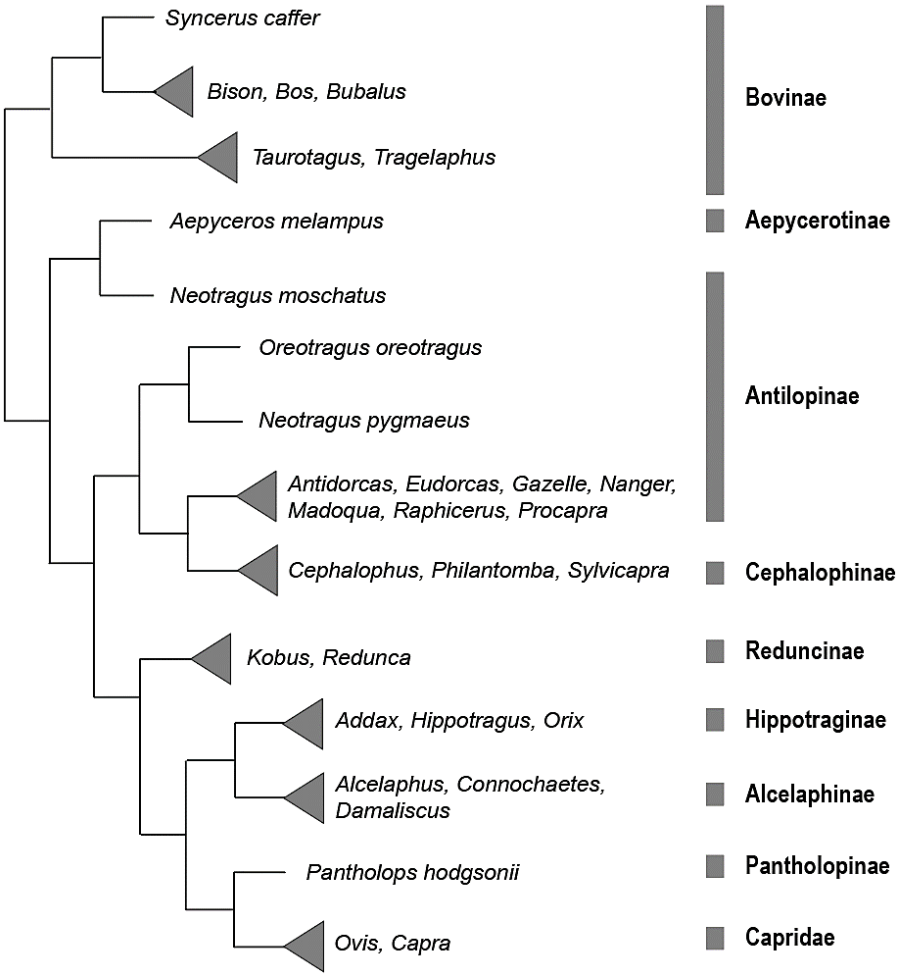


Figure 1. 7 Schematic cladogram showing intra-familial phylogenetic relationships within Bovidae family, adapted from Chen et al. (2019). Nodes are identified according to main genera composing them, and on the right is shown the corresponding sub-family.

1.2.3 African expansion

The first bovid-like fossil belongs to the extinct genus *Eotragus* and dates back to the Miocene, around 23 Mya (Vrba 1996; Bibi 2009). Because it was found in both France and sub-Saharan Africa, the common ancestor to all African antelopes is attributed to a species which either, extended its distribution, or was capable of long-distance migrations between continents (Matthee and Davis 2001; Bibi 2011; Moorley and Kingdon 2013). Based on the fossil record, it is also likely that Antilopinae, Reduncinae, and Capridae originated in Eurasia and subsequently expanded into Africa, around 10 to 14 Mya (Vrba 1996; Fernández and Vrba 2005). The earliest antelope-like fossil is from a member of the Antilopinae (*Pseudoeotragus seegrabensis*) found in Austria, estimated to be 17 to 18 million years. However, fossils attributed to the current *Gazelle* genus date back to the late mid-Miocene, around 14 Mya, across Kenya (Bibi 2009). This is coincident with the estimated divergence time for the Antilopinae node across the nuclear and mitochondrial genome of about 14 Mya (95% CI of 13.6-15.8 Mya; Bibi 2013; Chen et al. 2019). The oldest African records of Reduncinae were found in Chad and Kenya, and are dated to 8 Mya (Bibi 2009), which is consistent with an estimated divergence time of 7.1 Mya, from mitogenomic analysis (95% CI 6.3-8.0 Mya; Bibi 2013). The nuclear genome phylogeny, however, suggests an older ancestor for this sub-family, around 13 Mya (Chen et al. 2019). This period between the Miocene-Pliocene boundary, corresponds further with the earliest fossil records for Alcelaphinae (Kenya, 6.5-7.5 Mya), Hippotraginae (Chad, 7 Mya), and Tragelaphinae (Kenya, 5.7-6.7 Mya) (Bibi 2009).

Subsequent divisions within each sub-family and across genera occurred throughout the Pliocene (~2.5 to 5 Mya) and Pleistocene (~100 kya to 2.5 Mya) periods (Bibi 2013). The mid-Pliocene African climatic shift towards cooler temperatures and increased aridity, corresponds with widespread fossil record of grazing mammals and adaptation towards grasslands and savannas (Vrba 1996; Dupont, 2011; Bibi and Kiessling 2015). Across several paleontological deposits in East and South Africa, the characterization of faunal communities, through both species abundance and dietary isotopic measures, demonstrates an elevated turnover rate between 2 and 3 Mya (Bobe and Eck 2001; Bobe and Behrensmeyer 2004; Kingston and Harrison 2007; Faith and Behrensmeyer 2013). Analysis of bovid fossils along a geological timeline clearly shows a shift in species abundance from primarily browsers, associated with closed and wetter habitats (e.g Reduncinae), towards the ones that have a grazing nature, occurring mostly across more open and drier conditions (e.g Hippotraginae). Modern Hippotraginae species, for example, are primarily obligate grazers, distributed across desert grasslands and open woodlands. However, dietary isotopic measures across the Laetoli deposits, in northern

Tanzania, indicate that these species might have been mixed feeders or even primarily browsers during the Pliocene (between 3.5 and 3.8 Mya) (Kingston and Harrison 2007). Records at the same site from 2.5 to 2.7 Mya, show a tendency towards mixed foraging strategies, not only across Hippotraginae, but generally for bovid specimens. Studies like this show that, even though bovid evolution during this period can be largely attributed to a global climate shift, local environmental changes can shed light on complex biotic interactions at smaller scales (Bibi and Kiessling 2015).

1.2.4 Pleistocene refugia

As previously stated, Pleistocene climatic and vegetational changes shifted the range and variety of existing biogeographical regions across sub-Saharan Africa, and therefore, the spatial distribution of biological groups. How these climatic cycles influenced range expansions of taxa became a topic of interest (Hewitt 1996, 1999, 2000; Taberlet et al. 1998). The onset of phylogeography, by combining phylogenetic data on extant taxa with geographic and climatic history, allowed us to further understand how past events shaped current spatial patterns and genetic diversity (Avise 2009). Even though most studies were biased towards temperate biota across higher latitudes (Beheregaray 2008), several were performed across African savanna mammals (e.g. elephant; Nyakaana, Arctander and Siegismund 2002), including bovine (e.g. buffalo; van Hooft, Groen and Prins 2002) and antelope species (e.g. hartebeest, topi and wildebeest; Arctander, Johansen and Coutellec-Vreto 1999; kob antelope; Birungi and Arctander 2000; impala and greater kudu; Nersting and Arctander 2001).

By performing a comparative analysis across several phylogeographic studies on taxa with overlapping distributional ranges, it is possible to infer common patterns of spatial genetic discontinuities (Bermingham and Moritz 1998; Arbogast and Kenagy 2001). Hewitt (2004) compiled several phylogeographic studies on African savanna mammals to get some insight on how Pleistocene climatic cycles influenced their population structure, divergence, and diversity. Savanna habitats were prevalent across the continent throughout mid-Pleistocene, with high abundance around 1.2 and 1.7 Mya, and increasing prevalence towards late-Pleistocene, from 600 kya (Hewitt 2004). Also according to the author, divisions at the species level seem to be very recent and occurred mostly in the last 400 kya. Based on divergence level between clades, Hewitt (2004) identified three main areas across sub-Saharan Africa that consistently encompassed major divergences within species: West, East, and South (Figure 1.8). Phylogenetic analyses, patterns of population structure, and diversity measures, indicated that these three main regions most likely acted as refugia for savanna-adapted

taxa during moist pluvials. from where surviving populations would expand during subsequent drier interpluvials (Hewitt 2004). During these favourable periods, several re-colonisations between these main regions would occur, creating contact zones of previously isolated populations, with increased diversity (Hewitt 2001, 2011).

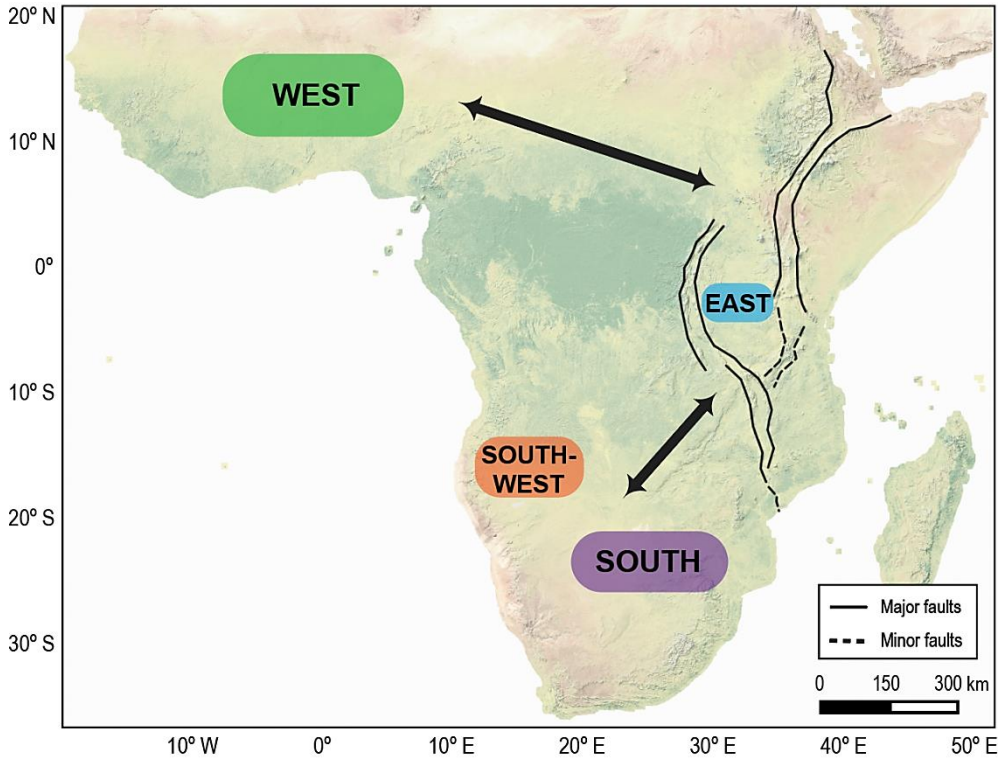


Figure 1. 8 Geographic location of main Pleistocene pluvial refuge areas for African mammals, adapted from Hewitt (2004) and Lorenzen et al. (2012): West (in green); South (in purple); South-West (in red); as well as East (in blue) acting as contact zone between the former. Black double-sided arrows depict range expansion movements during interpluvials. Also represented is the East African Rift System – EARS, adapted from Chorowicz (2005). Major faults are depicted in black curves, representing the eastern and western branches, as well as minor faults, in black dotted curves, namely the Eastern Arc Mountains – EAM. Land cover is represented from forests (in green) to mountains (in brown), with shaded relief, depicting higher altitude.

The continued publication of phylogeographic studies on African mammals made this pattern of Pleistocene refugia more apparent, especially among ungulate species strongly associated with the savanna biome, such as the giraffe (Brown et al. 2007), the roan antelope (Matthee and Robinson 1999; Alpers et al. 2004), and the sable antelope (Pitra et al. 2002, 2006). By combining their own work with available data on these species, Lorenzen and co-workers (2012) performed a robust comparative phylogeographic analysis. Across 19 ungulate species, 13 of which were antelopes, they validated Hewitt's proposal on Pleistocene refugia across West and South Africa, suggesting a second refugium in Southern Africa, centred in the southwest, for certain species (e.g. the sable antelope). They also suggested that East Africa acted mainly as

a contact zone during interpluvial range expansions, presenting itself as a mosaic of smaller refugia, with higher overall diversity levels. The complex pattern in East Africa was mostly driven by high species turnover, exacerbated by habitat instability associated with the continued volcanic and tectonic activity of the EARS (Trauth et al. 2009; Moorley and Kingdon 2013). The Pleistocene refuge theory thus postulates that the maintenance of non-forestry refugia during pluvials enabled the persistence of species through repeated glacial cycles. These patches of suitable habitat promoted the divergence between refuge populations during periods of isolation, as well as the accumulation of genetic diversity due to species' persistence. Range expansion rates during interpluvials and the presence of geomorphological barriers, dictated the position of contact zones, and ultimately shaped the spatial patterns that we see today across African taxa, namely savanna-adapted antelope species (Lorenzen et al. 2012).

1.3 *Hippotragus* Genus

Hippotragus HARRIS 1838 belongs to the Hippotraginae SUNDEVALL 1845 sub-family, also known as the horse-like antelopes due to their grazing nature and body proportions. Is an African endemic genus, including three species of antelopes: the extinct bluebuck (*H. leucophaeus*); and two extant species, the roan (*H. equinus*) and the sable antelope (*H. niger*) (Figure 1.9). As previously stated, unlike remaining antelopes within the Hippotraginae sub-family, *Hippotragus* species are highly associated with the savanna biome and, therefore, share many life-history characteristics (Chardonnet and Crosmary 2013; Estes 2013).

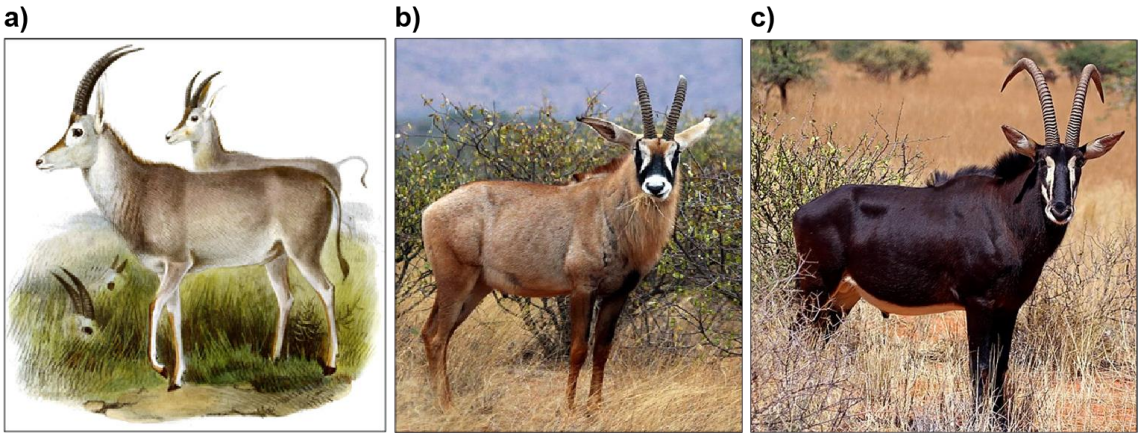


Figure 1. 9 Members of the *Hippotragus* genus. a) male and female bluebuck (*H. leucophaeus*), illustration by Smit and Wolf, retrieved from Sclater and Thomas (1899); b) male roan antelope (*H. equinus*) photo by Sharp, 2014 (https://www.sharpphotography.co.uk/); c) male sable antelope (*H. niger*), photo by Sharp, 2014 (https://www.sharpphotography.co.uk/).

1.3.1 Species description

The bluebuck, or blue antelope (*Hippotragus leucophaeus* PALLAS 1766) was the smallest of its congeneric species. It was endemic to South-Africa, with a restricted distribution to the southwestern Cape coast (Klein 1974). It was the first African mammal to become extinct during historical times, with the last individual being hunted by European settlers, around 1800 (Klein 1974). However, by the time it was described in 1766 by Pallas, it was already considered uncommon. Habitat fragmentation due to long-term climate change during the early Holocene, together with habitat specialization, might have reduced the population's genetic resilience before the 17th century. Continued habitat loss and overhunting during the 1800's would ultimately lead this species towards extinction (Kerley et al. 2009; Faith and Thompson 2013).

The roan antelope (*Hippotragus equinus* GEOFFROY 1803) is the biggest among its congeneric species and one of the largest among other African antelopes. From head to tail, they can measure between 190 cm and 240 cm, and weight between 220 kg to 300 kg. Its name comes from a general reddish-brown coat colour, although having a black face, which is lighter in females, with characteristic white patches around the eyes and mouth. They have characteristic floppy years. Both sexes have ringed horns that arch slightly backwards and are bigger in males (between 55 and 100 cm) (Chardonnet and Crosmary 2013). It is a gregarious animal with strong social structure, living up to 17 years in the wild. Dominant roan bulls usually defend a herd of six to fifteen females plus their offspring, from other bachelor male groups. Territories can range from 250 hectares to 1,000 hectares, at normal density, and herd composition can change daily. Breeding can occur throughout the year, without a particular season, and gestation period takes about nine months. After three to four weeks of hiding, calves join other juveniles in associated maternal herds, where they are weaned until six months of age. Females become sexually active within two years, and by that time, sub-adult males are forced out of the territory by the dominant bull. These evicted males disperse and join in bachelor groups, until they reach maturity around six years of age. Males fight among themselves for a dominant position within the herd, using their horns; although brutal, these fights are rarely mortal (Chardonnet and Crosmary 2013). The species inhabits shrubland and grassland savannas across sub-Saharan Africa. At a finer scale, it has a patchy and discontinuous distribution, due to specific habitat requirements of moist soils and constant availability of water, grazing predominantly on open land, with medium to tall grass (Codron et al. 2007; Chardonnet and Crosmary 2013). They occur widely across sub-Saharan Africa, from West and Central Africa, where it is more common, to

East and Southern Africa, where it is now rare (East and IUCN SSC Antelope Specialist Group 1999; IUCN SSC Antelope Specialist Group 2017a) (Figure 1.10).

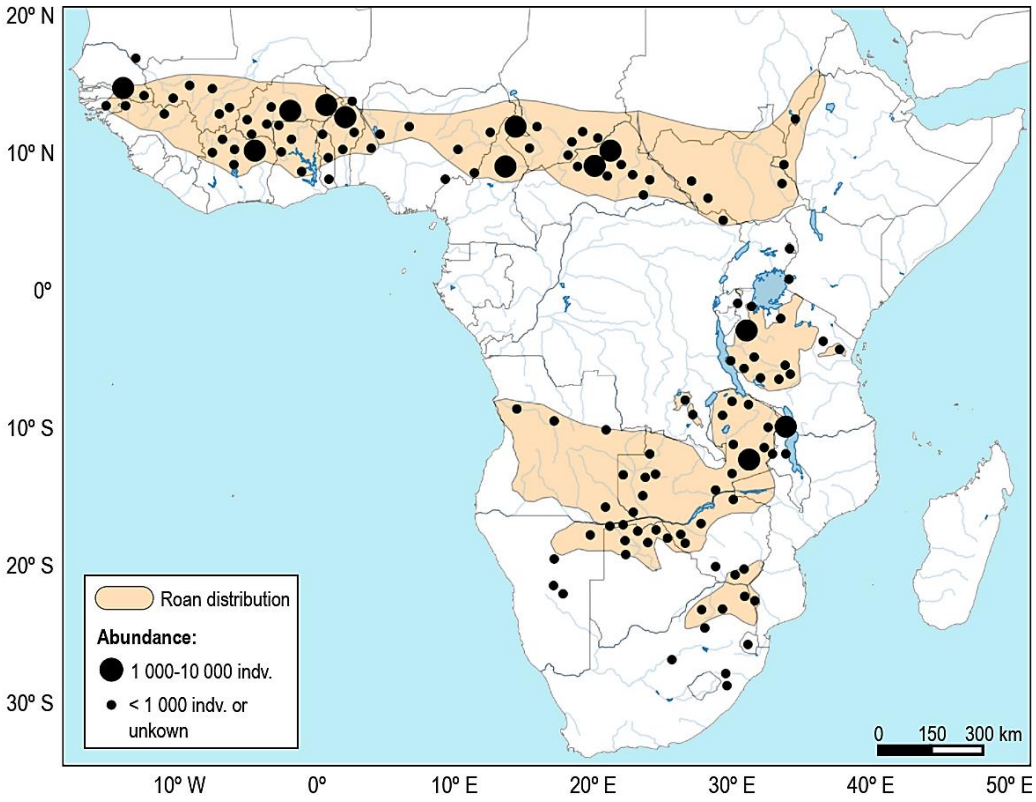


Figure 1. 10 Distribution of the roan antelope (*H. equinus*) across Africa, in orange, according to the IUCN (2017a) and population abundance, represented by black circles, according to East and IUCN SSC Antelope Specialist Group report (1999). Bigger circles indicate abundance numbers between 1,000 and 10,000 individuals; smaller circles indicate abundance lower than 1,000 individuals, or unknown. Abundance points outside current specie's distributional range, due to population contractions up until the early 2000's.

Six subspecies are described for the roan antelope, with geographical correspondence: *H. e. koba* GRAY 1872, in the northwest, from Senegal to Nigeria; *H. e. charicus* SCHWARZ 1913, in North-central Africa, from Nigeria to Central African Republic (CAR); *H. e. bakeri* HEUGLIN 1863, to the northeast, across Sudan, Ethiopia, and northern Uganda; *H. e. langheldi* MATSCHIE 1898, to the east, through southern Uganda, Kenya, Tanzania, Rwanda, and Burundi ; *H. e. cottoni* DOLMAN AND BURLACE 1928, in South-central Africa, across northern Malawi, Zambia, southern Mozambique, and Angola; and *H. e. equinus* DESMAREST 1804, in the south, from southern Mozambique towards Zimbabwe, Botswana, Namibia, and South-Africa (Ansell 1972).

The sable antelope (*Hippotragus niger* HARRIS 1838) is characterized by its strong sexual dimorphism, with bigger sized and darker-coloured males. From head to tail, males measure between 190 cm and 250 cm, and weight on average 235 kg. Females and juveniles present chestnut to dark brown coat colour, whereas males begin

darkening and turn ebony black around three years of age. They present the same black faces with white stripes as roan antelopes, as well as curved-shaped horns that reach between 80 cm, in females and 125 cm, in males (Estes 2013). As the roan, it is a sociable and territorial antelope, living up to 17 years in the wild. Sable bulls are strictly territorial whereas females and offspring form hierarchical mobile groups, ranging between 10 to 30 individuals, across 10 to 50 hectares. Females usually have one oestrous cycle per breeding season, from May to July, with mating season occurring mostly in June. After nine months gestation, births happen at the end of the rainy season, when long grass is available for grazing. As with roan antelopes, calves are hidden during the first three weeks and then join the maternal group, being weaned for six months, until the end of the dry season. Females become sexually active around two and a half years of age, and live in maternal herds, with subordinate juvenile males. By the time males become three to four years old, they are forced out of the maternal herd to form bachelor groups until they reach maturity, around five years of age. During breeding season, sable bulls fight for herd dominance, using their horns; as in roan antelopes, although brutal, these fights are rarely mortal (Estes 2013). The species inhabits shrubland and grassland savannas, being strongly associated with the miombo (*Brachystegia* spp.) woodland zone and constant availability of water. During the wet season it occurs mostly in mesic woodlands with medium to tall foliage; in the dry season it occupies grasslands, feeding on smaller and greener herbage that emerges after annual fires (Codron et al. 2007; Parrini and Owen-Smith 2010; Estes 2013). It has a more restricted distribution than roan, occurring in East and Southern Africa, where both species are sympatric, with an isolated population in central Angola, evaluated as critically endangered (East and IUCN SSC Antelope Specialist Group, 1999; IUCN SSC Antelope Specialist Group 2008, 2017b) (Figure 1.11).

Four subspecies are described for the sable antelope, with geographical correspondence: *H. n. roosevelti* HELLER 1910, in the east, across southeast Kenya, Tanzania, and Mozambique; *H. n. kirkii* GRAY 1872, in southcentral Africa, through southern DRC and Zambia; *H. n. niger* HARRIS 1838, in the south, from western Mozambique to Namibia; and *H. n. variani* THOMAS 1916, across central Angola (Ansell 1972).

1.3.2 Conservation and management

In 1998, the IUCN/SSC Antelope Specialist Group (ASG) assembled a report on the conservation status of African antelopes and management actions for the continued survival of these taxa (East and IUCN SSC Antelope Specialist Group 1999).

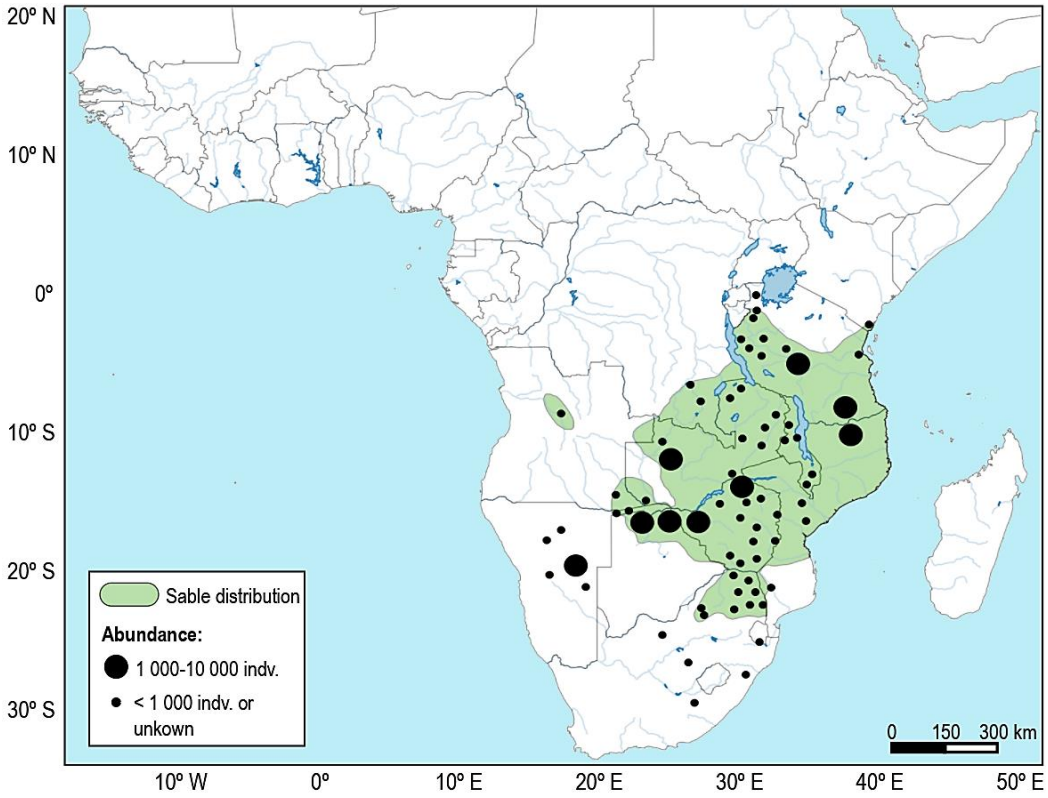


Figure 1. 11 Distribution of the sable antelope (*H. niger*) across Africa, in green, according to the IUCN (2017b) and population abundance, represented by black circles, according to East and IUCN SSC Antelope Specialist Group report (1999). Bigger circles indicate abundance numbers between 1,000 and 10,000 individuals; smaller circles indicate abundance lower than 1,000 individuals, or unknown. Abundance points outside current specie's distributional range, due to population contractions up until the early 2000's.

Main threats, identified as habitat loss, hunting pressure, and epizootic diseases, were related to human population growth and consequent expansion of human activities. Therefore, the main goal was to preserve as many areas possible of suitable habitat that would maintain as many viable populations as possible, while considering human development needs. This includes a range of land-use options, from protected national parks and reserves, with complete human restrictions, to game reserves, where sustainable use of wildlife, as well as significant generation of a revenue, are allowed.

The Regional Action Plan for Antelope Conservation quantitatively identified key areas for the conservation of antelope communities, across sub-Saharan Africa (see East and IUCN SSC Antelope Specialist Group 1999). The information collected indicates that wildlife is depleted or severely depleted across more than half of such areas. The worst status region is West Africa (covering Mauritania through Nigeria), with an average of 80% affected areas, and the least worst status region is Southern and South-central Africa (from Angola to South-Africa), with an average of 33% affected areas. Based on distribution maps of each species, ASG recorded species abundances across sub-Saharan Africa, from 1987 to 1998. However, accurate definition of the

precise distribution of most African antelopes is exceedingly difficult, especially considering current reduction and fragmentation of once large and continuous distributional ranges. Subspecies are generally ignored for conservation action plans, due to uncertain scientific validity and precise distribution of such entities. Exceptions only occur for on-field recognizable characters based on morphological, behavioural, and/or geographical distinctiveness. The only recognizable subspecies across *Hippotragus* for conservation purposes is the Angolan Giant sable (*H. n. vairiani*) (IUCN SSC Antelope Specialist Group 2008).

Based on the ASG report, both roan and sable antelopes were considered as "Conservation Dependent" species, whose continued survival was reliant on active conservation measures and breeding programmes. The roan antelope global population trend was set as decreasing, eliminated from large areas of its once wider range (Figure 1.10). At that time, in 1998, total population was estimated to be around 76,000 individuals, with largest populations – accounting 60% of total individuals – living only within protected areas. Outside protected areas, a smaller part of roan populations occur on private land in Southern Africa, where population trend is increasing (East and IUCN SSC Antelope Specialist Group 1999; IUCN SSC Antelope Specialist Group 2017a). Conservation measures to prevent continued roan antelope's decline included active protection and management across existing populations. An example of such measures is the National Recovery Plan in Kenya, implemented to re-establish a viable population in Ruma National Park (Kimanzi and Wanyingi 2014). Today, Ruma National Park is the only current area in Kenya where the roan sustains a viable population, after becoming locally extinct in Masai Mara National Reserve (IUCN SSC Antelope Specialist Group 2017a). Although efforts have been made, the roan antelope is also considered as locally extinct in Serengeti National Park, in Tanzania, and locally endangered across protected areas in South-Africa, namely in Kruger National Park, Nylsvley Nature Reserve and Percy Fyfe Nature Reserve (Grobler and Nel 1996; Harrington et al. 1999; Dörgeloh 2001; Kröger and Rogers 2005), with decreasing global populational trends and current estimated size between 50,000 and 60,000 wild individuals (IUCN SSC Antelope Specialist Group 2017a).

According to the ASG report (East and IUCN SSC Antelope Specialist Group 1999) sable antelopes were also eliminated from large areas of its once wider range, however global population trend was set as generally stable across larger populations in East and Southern Africa, in Tanzania, Zambia, Malawi, Botswana, Zimbabwe, and South-Africa (Figure 1.11). Today, total population is estimated to be between 50,000 and 60,000 individuals in the wild (IUCN SSC Antelope Specialist Group 2017b). Outside protected areas, increasing numbers of sable antelopes occur in private land in Namibia,

Zimbabwe, and South-Africa, as well as captive populations across North America. The only exception to this trend is the isolated population of the Angolan Giant sable (*H. n. variani*), considered as “Critically Endangered” by the IUCN Red List of Threatened Species (IUCN SSC Antelope Specialist Group 2008). During Angolan Civil War (1975-2002), this population suffered a severe decline due to hunting pressure, decreasing from an estimated 2,000 to 3,000 individuals before 1970, to around 300 in 2007 (IUCN SSC Antelope Specialist Group 2008). The population was feared extinct until its rediscovery in 2006 (Pitra et al. 2006), which led to the development of an *in situ* Conservation Plan to ensure suitable habitat and management of surviving individuals, from both an ecological and genetic perspective (Vaz Pinto 2009; Estes 2013; Vaz Pinto et al. 2016). Currently, 100 Giant sable individuals are estimated to live within remaining areas of Cangandala National Park and Luando Reserve (IUCN SSC Antelope Specialist Group 2008). In recent years, increasing numbers of sable antelopes occurs in private land in Zambia, Namibia, Zimbabwe, and South-Africa, as well as captive populations across North America, mainly because it has become one of the most highly prized trophy hunting species in Africa (Bothma and van Royen 2005; Lindsey et al. 2013; Piltz, Sorensen and Ferrie 2016).

Inbreeding and outbreeding depression risks are well documented within endangered species (Edmands 2006). Interspecific hybridization between the roan and sable antelope can occur under exceptional circumstances, as the one reported for the Angolan Giant sable (Vaz Pinto et al. 2016). However, intraspecific hybridization may be of more generalized concern for these species, owing to animal translocation across different subspecies ranges, as documented recently for roan antelopes (van Wyk et al. 2019). Contemporary conservation studies rely on defining units of management based on both genetic and ecological distinctiveness of a species, subspecies, or population, inferred by phylogenetic history. They are defined as evolutionary significant units (ESUs) or management units (MUs), either representing a slower or faster isolation response, respectively (Moritz 1994, 1999). Genomic data allows for the inference of conservation units (CUs), integrating adaptive genetic variation (Funk et al. 2012). From a conservation and management perspective, it is of extreme importance to define such units and prevent further depuration of genetic diversity.

1.3.3 Evolutionary history

Several phylogeographic studies have been assessing levels of population structure, diversity, and differentiation across both roan and sable antelopes, while attesting for

subspecific validity. Observed differentiation patterns can be associated with specific evolutionary histories, shaped by both intrinsic and extrinsic factors.

The first attempt to study the roan antelope using molecular data was by Matthee and Robinson (1999), analysing a fragment of the mitochondrial DNA (mtDNA) control region. They included 13 individuals covering the range of all, but two recognized subspecies, the north-central *charicus* and north-eastern *bakeri*. Phylogenetic analysis retrieved four mtDNA lineages, supporting the division between northwest *koba* (Benin), eastern *langheldi* (north Tanzania and Kenya), south-central *cottoni* (south Tanzania, Malawi, and Zambia) and southern *equinus* (Namibia and South-Africa) subspecies. These associations were related to vicariance or the effects of isolation by distance, and a clear division between individuals in East and Southern Africa. However, the low number of individuals and the use of a single mtDNA fragment is limiting for evolutionary inferences. In a subsequent study, Alpers and co-workers (2004) extended the analysis of the same mtDNA fragment plus eight microsatellite loci, to more than one hundred individuals, covering the same four subspecies but with a more comprehensive sampling, and adding one sample from the subspecies *charicus* from Cameroon. Based on phylogeny and individual clustering, these authors proposed the presence of two ESUs corresponding to the northwest *koba*, and remaining distribution range, respectively. The sample representing the subspecies *charicus* from Cameroon was not strongly associated with any of the analysed subspecies and its phylogenetic position was left unresolved. Additionally, these authors proposed two refugial areas for the species during the Pleistocene climatic cycles, located in the West and East of Africa, respectively. The West refuge, associated with *H. e. koba*, was proposed to be specifically located around Ghana, while the East refuge comprised the remaining subspecies, with the most likely location of the refugial population around the northeast. Both studies added relevant information on roan's genetic diversity and evolutionary history, but the poor geographic sampling, particularly across north-central and northeast distribution of the species, left uncertainty on the number of genetic clusters across the species distributional range and have not investigated the time of their divergence (Figure 1.12). Considering previous studies on the roan antelope, Lorenzen et al. (2012) associated the highest levels of gene flow for the eastern samples to a possible recent admix and a mosaic of several other possible refugia in the area.

Matthee and Robinson (1999) also analysed the mtDNA control region for 15 sable individuals, covering the four described subspecies. Phylogenetic analysis retrieved two mtDNA lineages, dividing specimens from west Tanzania from the remaining ones. This differentiation was related to a geographic barrier, including vegetational differences and rifting, that disrupted gene flow between both lineages. Soon after, Pitra and co-workers

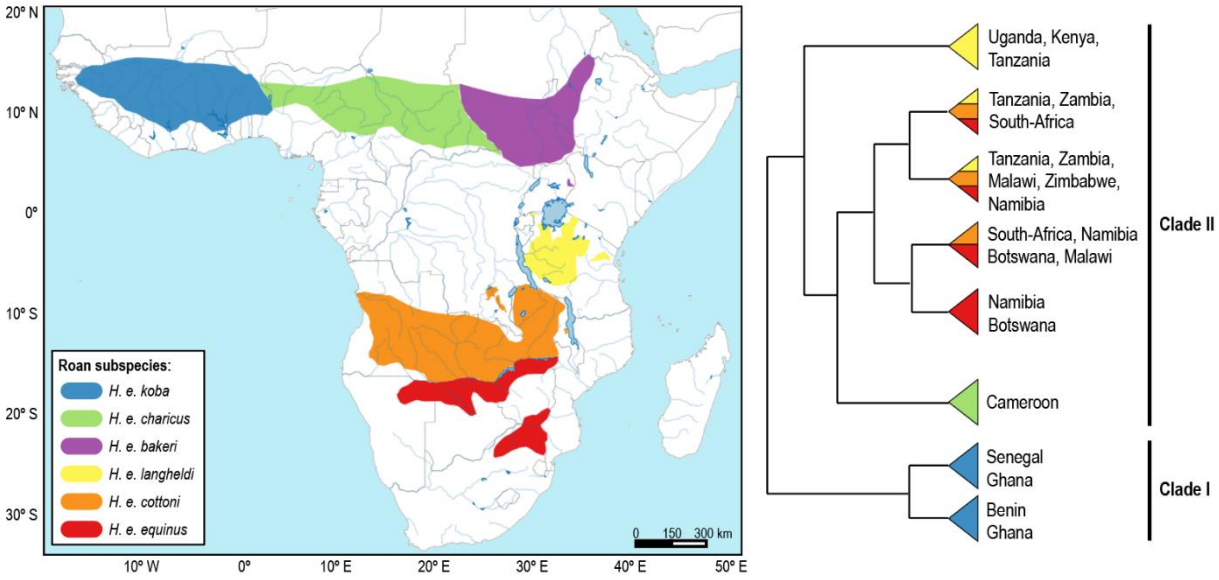


Figure 1. 12 The **r**oan antelope intraspecific analysis. On the left, subspecies geographical distribution, according to Ansell (1972) and current species distribution (IUCN SSC Antelope Specialist Group 2017a). On the right, a schematic cladogram showing intraspecific phylogenetic relationships using mitochondrial data, adapted from Alpers et al. (2004); clade identification following the ones from the authors. Each colour corresponds to a subspecies, according to the legend.

(2002) supported this observation by combining mtDNA control region and cytochrome b sequences of 95 individuals from several localities, across East and Southern Africa. They identified three mtDNA lineages, one in west Tanzania, another in the East, and a third one in the Southern Africa, but also including some specimens from west Tanzania. The observed sympatry of the two highly divergent mtDNA lineages in west Tanzania was explained by an ancient unidirectional outbreeding event. Sable antelope's evolutionary history was explained taking into account several historical events, including allopatric fragmentation between west Tanzania and the remaining populations, consistent with the discontinuous distribution of the miombo habitats around the mountain circles in East Africa, as well as alternated periods of short to long-distance dispersal between individuals in west Tanzania and those in eastern and southern ranges, respectively. Later, by including samples of the Angolan Giant sable in their previous dataset, Pitra et al. (2006) observed a close relationship between Angolan haplotypes and those observed in a few individuals from western Tanzania. Once again, these patterns were associated with long-distance dispersal between both regions, from a common source population. With this result, the legitimacy of the Giant sable as a separate clade was left unresolved. The validity of the *variani* subspecies was further investigated by Jansen van Vuuren and co-workers (2010), specifically its genetic distinction from geographically closer and phenotypically similar, particularly with respect to facial markings, individuals living across western Zambia. The authors concluded that

despite their morphological resemblance, significant genetic differences underpin these two evolutionary lineages.

Using whole-genome mitochondrial sequencing for more than 200 contemporary and historic samples, across the sable antelope's entire distributional range, Rocha (2014) unveiled the species' maternal evolutionary history. Four main lineages were identified, including the divergent lineage identified in previous studies in west Tanzania – which was assumed to be more closely related to *kirkii* subspecies, according to Ansell (1972), and later named *relict* by Vaz Pinto (2018); an eastern lineage present in Kenya and north Mozambique; southern lineage comprising individuals from south Mozambique, Zambia, Zimbabwe, Botswana, Namibia, and South-Africa; and a central lineage present in Angola, west Tanzania, and Malawi. High divergence levels (~3%) found between the relict and the remaining lineages, and associated vicariant event estimated around 1.4 Mya, led to the hypothesizes that the relic lineage was a ghost lineage of an extinct population that survived, as a result of an ancient introgression event. The origin of this lineage, dated to the early Pleistocene period (~1 to 2.5 Mya), was related to the increased tectonic activity of the EARS, probably contributing to the isolation of a population in the Tanzanian plateau, from the remaining individuals in the East. The second vicariant event, dated to 350 kya, subsequently separated the eastern lineage, delimited by the EARS and the EAM, across central Tanzania. Central and southern lineages split later, around 200 kya. The presence of the central lineage across individuals in west Tanzania, Malawi, and Angola was explained by a dispersal corridor through Congo. Individuals found today in west Tanzania carry both the relict and central lineage, which clarifies the proximity found in previous studies between *variani*, *niger* and west Tanzanian samples (Pitra et al. 2002, 2006; Jansen van Vuuren et al. 2010). Further divisions observed within eastern and southern lineages were dated around 150 kya and coincide with the Pleistocene dry cold period MIS 6. These divergences were explained by vegetation change in both regions, caused by climatic transition, with subsequent reconnection by range expansion, once conditions improved. The split within haplogroups of the southern lineage occurred after 100 kya and was associated with the hydrological changes the Zambezi River experienced, separating individuals in the northern from those in the southern bank of the river.

The first study on the sable antelope's evolutionary history using nuclear data was by Vaz Pinto (2018), genotyping 400 individuals for 57 polymorphic microsatellites previously developed (Vaz Pinto et al. 2015). Using clustering analysis, five geographically differentiated groups were obtained and named as follows: west Tanzanian, in the Tanzania plateau; eastern, distributed from Kenya to central Mozambique, Malawi, and eastern Zambia; southern, located south of the Zambezi

River; Zambian, across Zambia to the east and north of the Zambezi, including south DRC; and Angolan, across Cangandala National Park and Luando Reserve in Angola. Angolan group registered the lowest levels of diversity and highest levels of differentiation, in agreement with its restricted location and recent population bottleneck. These five groups are consistent with previous study based on mitogenomes (Rocha 2014) and their geographical delimitation supports vicariant events resulting from the same geographical features, such as the EARS, EAM and the Zambezi River, as well as Pleistocene climatic fluctuations (Figure 1.13).

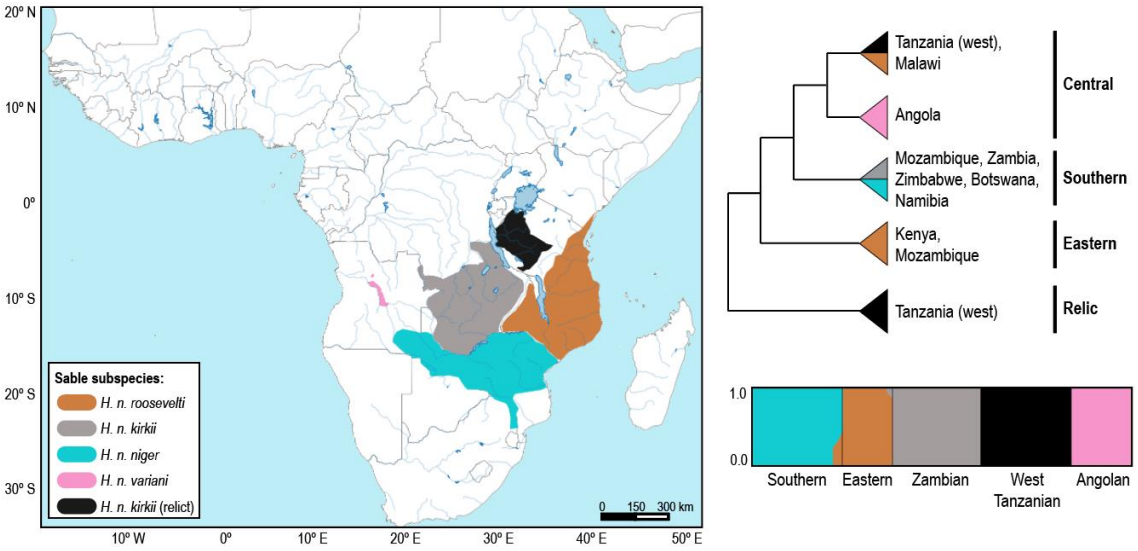


Figure 1. 13 The sable antelope intraspecific analysis. On the left, subspecies geographical distribution, according to Ansell (1972) and current species distribution, adapted from IUCN SSC Antelope Specialist Group (2008, 2017b). On the right, a schematic cladogram showing intraspecific phylogenetic relationships using whole-mitochondrial data, adapted from Rocha (2014), as well as clustering analysis showing the assignment proportions of each individual using microsatellite loci, adapted from Vaz Pinto (2018); mtDNA lineages and nuclear groups identification following the ones from the authors. Each colour corresponds to a subspecies, according to the legend.

Based on the studies of Matthee and Robinson (1999) and Pitra et al. (2002, 2006), Lorenzen et al. (2012) proposed three main refugial areas for the sable antelope during the Pleistocene climatic cycles, located in the East, South, and Southwest Africa, respectively. The East refuge was associated with the highest divergence levels calculated for the population in west Tanzania, and therefore, supposed to be the oldest population. The South refuge comprised remaining sable antelopes' distributional range, except for the isolated *H. n. variani*, which was associated with a Southwest refugial area, probably deriving from the South refuge, with subsequent independent evolution. Based on later studies from Rocha (2014) and Vaz Pinto (2018), the reasons for the East and Southwest refugia were not validated.

1.4 Molecular markers: from population genetics to population genomics

A DNA sample contains the signature of the organism from which it was taken, at both population and species level. This is because the DNA molecule is transmitted from one generation to another, accumulating differences that occur in populations over time. The discipline that studies these differences is called population genetics, and when a population study involves a huge amount of the entire genetic information of an organism, is generally referred to as population genomics (Luikart et al. 2003).

1.4.1 Types of molecular markers

An eukaryotic cell contains different types of DNA, namely the nuclear DNA (nDNA), the mitochondrial DNA (mtDNA), and in the case of plants, the chloroplast DNA. Animal cells contains several hundreds to thousands of mitochondria, each with two to ten copies of DNA, and therefore, mtDNA is potentially more easily available when compared to nuclear DNA, with only one copy per cell (Xia et al. 2009). MtDNA and nDNA have been extensively used for population genetic studies, each with their own strengths and disadvantages, due to their inheritance nature, distribution across the genome, and rate of evolution (Sunnucks 2000; Zhang and Hewitt 2003).

Mitochondrial DNA has a matrilineal inheritance mode, with little or no recombination (Rokas, Ladoukakis and Zouros 2003). Vertebrate mtDNA contains 13 protein-coding genes, 22 transfer RNAs (tRNAs), and two ribosomal RNAs (rRNA) (Broughton, Milam and Roe 2001). In addition, mtDNA has a non-coding region, termed control region due to its role in replication and transmission of mtDNA molecules (Sigurðardóttir et al. 2000). The insufficient repair rate for mtDNA causes high rates of nucleotide substitution, compared with most regions of the nuclear genome (more than 10 times higher for vertebrates) (Allio et al. 2017). Factors deriving from intergenomic co-adaptation can cause variation in evolutionary rates across mtDNA loci, as well as variations of the same loci across taxa (Hagelberg 2003; Galtier et al. 2009).

Microsatellites or simple sequence repeated (SSR) loci are found throughout the nuclear genome. About 10^4 to 10^5 microsatellite loci are held in the genome, and because of that, they can be isolated from virtually any eukaryotic species of interest (Charlesworth, Sniegowski and Stephan 1994). They can range in length from one to six nucleotides in length, which are tandemly repeated from five to a hundred times (Richard, Kerrest and Dujon 2008). Microsatellites exhibit high mutation rates, generally in the order of 10^{-3} to 10^{-4} per allele per generation, which generates a large amount of allelic variation in the number of repeats (Rienzo et al. 1994; Estoup, Jarne and Cornuet 2002),

and because of their highly polymorphic nature have been for decades the marker of choice in population genetics studies (Bruford and Wayne 1993; Ellegren 2004; Mittal and Dubey 2009). Being relatively easy to genotype, microsatellites are also suitable for molecular analyses with limited amount of DNA or when the DNA quality is poor (e.g. non-invasive samples), since it is PCR-based (Morin et al. 2001). Microsatellites have a co-dominant Mendelian inheritance and are mainly located at non-coding regions. Nevertheless, their diverse distribution, including at non-neutral regions, determines different selective pressures among loci and taxa, with great variation on the mode and *tempo* of evolution (Li et al. 2002). The major practical constraints in microsatellite genotyping are precisely related to high levels of polymorphism, which hinders the design of common primers for non-related species (Rubinsztein et al. 1995; Zane, Bargelloni and Patarnello 2002), and also leads to amplification errors (Schlötterer and Tautz 1992; Viguera, Canceill and Ehrlich 2001).

In 1996, Lander proposed the use of nuclear single nucleotide polymorphisms (SNPs) as a valid molecular marker in population genetics (Lander 1996), either being a single nucleotide substitution or single to multiple base insertions or deletions, collectively called indels. Nuclear DNA contains the most extensive genetic information of an individual, and when compared to both microsatellites and mtDNA, generally presents the slowest evolutionary rate, especially across mammals (Vignal et al. 2002; Wan et al. 2004). SNPs substitution rate, however, depends on the genome region, varying according to levels of recombination, coding or non-coding nature of the region, under neutral or selective pressures, among others (Yang and Nielsen, 1998; The International SNP Map Working Group et al. 2001; Nachman 2002; Hodgkinson and Eyre-Walker, 2011). What was a “mirage” by 1996, only reachable for the human genome, became a common molecular tool for every species nowadays, highly simplified by the advent of high throughput sequencing technologies (NGS), that turned possible to obtain large quantities of DNA sequences, from where thousands of SNPs can be extracted and analysed. The genome sequence of several model organisms, namely humans, has shown that SNPs are mostly found across neutral, non-coding regions (Venter et al. 2001).

1.4.2 Applications within population studies

The use of different molecular markers in population genetic studies has been exponential, and their studies related to population structure and evolutionary history has had a crescendo in accuracy and complexity (Schlötterer, 2004; Grover and Sharma 2016). The genetic signature that individuals carry in their genome allows to quantify the

relative importance of evolutionary factors, such as migration, genetic drift, or selection, on present-day populations (Chikhi and Bruford 2005).

The studies on taxa evolution and phylogenetics were highly boosted by the analysis of molecular data, for which the mitochondrial DNA had an early and significant contribution (Avise et al. 1987; Avise 1998). The different mutation rates across the several regions of the mtDNA molecule have soon been used to investigate evolutionary-related questions across different categorical levels (Simon et al. 1994). For example, the highly conserved region of 12S rRNA has been employed to illustrate old phylogenetic subjects, for instance across phyla and sub-phyla, whereas 16S rRNA has been more usually applied to investigate phylogenies of families and genera (Kuznetsova, Kholodova and Luschekina 2002). The mitochondrial protein-coding genes evolve much faster than the rRNA and therefore have been used as powerful markers for the inference of evolution history from families to species (Rebholz and Harley 1999; Jansen van Vuuren and Robinson 2001), whereas the two hypervariable domains of the control-region have been mainly used in intraspecific studies (Kholodova et al. 2006). Despite the high mutation rate of these regions, it is not enough to analyse individual differences nor very recent changes in population parameters. The major limitation in mtDNA is related to its inheritance nature, restricting the analysis of evolutionary events to the matrilineal lineage, and because of this, evolutionary and demographic inferences based on this genetic marker should consider the fact that its effective population size corresponds to a fourth of that of nuclear autosomal markers (Ballard and Whitlock 2004). Discordant patterns between mitochondrial and nuclear DNA are commonly found among taxa, and the most parsimonious choice often involves the combined use of both markers (Toews and Brelsford 2012). The advent of NGS expedited the process of obtaining whole mitochondrial genomes (mitogenomes) sequences, with the advantage of extending analyses to historic samples (Curole and Kocher 1999; Stiller et al. 2009; Knapp and Hofreiter 2010; Mamanova et al. 2010; Maricic, Whitten and Pääbo 2010; Dabney and Meyer 2012; Dabney et al. 2013). This enables to consider different rates of evolution across regions, increasing the phylogenetic resolution and estimation accuracy of divergence time intervals (Rokas and Carroll 2005; Cui et al. 2007; Yu et al. 2007; Hiendleder, Lewalski and Janke 2008; Botero-Castro et al. 2013; Espregueira Themudo et al. 2015).

Microsatellite data have been extensively used in population genetic studies, including among African antelope species (Eblate et al. 2011). The higher mutation rate of microsatellites allows to infer intraspecific and individual level genetic variation in contemporary events, with only a few set of loci (Pepin et al. 1995; Luikart and England 1999; Vaz Pinto et al. 2015). The statistical power of microsatellite markers, however,

depends on the degree of polymorphism of each locus, as well as the sample size, and if it represents the general variability across the studied taxa (Zhivotovsky and Feldmant 1995; Väli et al. 2008). NGS technologies have facilitated the development of species-specific microsatellites (Abdelkrim et al. 2009; Gardner et al. 2011), limiting problems of their usage related to ascertainment bias (see previous section).

By relying on thousands of SNPs, scattered across a genome, sequencing studies provide highly accurate inferences on population parameters like genetic structure (Liu et al. 2005), demography (Brumfield et al. 2003), and adaptation (Shastry 2002). The assembly of genomes across model species, which include many domesticated taxa, instigated the genomic era across non-model species. By using reference genomes to map against whole-genome sequencing of non-model species, it is now possible to increase accuracy in the understanding of all population genetics inferences, including the evolutionary history across natural populations (Luikart et al. 2003). However, third-generation sequencing is providing easy access to *de novo* genome assemblies of non-model species, allowing to not only identify SNPs across the genome, but also to characterize structural variation, and understand gene and protein function, making it a fundamental genomic tool for future research (Bleidorn 2016; Etherington et al. 2020).

1.5 Objectives and thesis organization

This thesis is focused on the evolutionary history of the roan antelope across its range in sub-Saharan Africa, as well as on the comparative genomic analyses between roan and sable antelopes. Using both genetic and genomic makers, we analyse several statistics at the intraspecific level, towards improved taxonomic clarification and phylogeographic context. We contextualize *Hippotragus* evolution within African's climate and geomorphological history, and how it has been shaping the phylogeographic patterns across antelope species, in particular, as well as remaining taxa, in general. Towards this goal, this thesis also presents the first review on genetic discontinuities across animals (vertebrate and invertebrate) and plants in sub-Saharan Africa, allowing for a comprehensive analysis of the most important geomorphologic features that have shaped the genetic diversity and differentiation patterns of savanna-adapted taxa, over time.

Specific objectives are as follow:

i) To review the genetic discontinuities across savanna-adapted taxa, by performing a comparative analysis on published scientific work;

ii) To assess the genetic diversity and population structure for roan antelope across its entire distributional range, combining modern and historical samples;

iii) To estimate divergence time intervals across roan antelope mitochondrial lineages and relate these to Pleistocene climatic fluctuations;

iv) To perform demographic analysis across identified populations for roan antelope, and contextualize with current trends;

v) To provide genomic resources for future research within the roan antelope, in particular, and remaining African antelopes, in general;

vi) To estimate genomic levels of population structure, diversity, and evolution for both the roan and the sable antelope, across their entire distributional range;

vii) To do a comparative genomic approach between both species and relate differential patterns with both intrinsic and extrinsic factors.

The thesis is organized in five main chapters. **Chapter 1** corresponds to the current one, with a *General Introduction* to the multiple subjects included in this thesis, divided into four sections: the first one regards sub-Sahara African biogeography and climatic history; the second, a summary on African antelope characteristics, their origins and consequent African expansion; then, is given an overview on *Hippotragus* genus, with a description upon extinct and extant species, and their conservation and evolutionary history; the fourth and final section contains a summary upon molecular markers and their development towards population studies.

Chapter 2, **Chapter 3,** and **Chapter 4** contain research studies upon several aspects regarding the thesis' objectives, presented in the form of scientific manuscripts, which are either published, submitted, or in preparation for future publication. In **Chapter 2** we provide a comparative analysis on sub-Saharan Africa genetic discontinuities, in the form of a review of published scientific work across several taxa. **Chapter 3** regards population analyses of the roan antelope and is divided into two research studies. *Study I* evaluates the roan antelope's genetic structure across its distributional range, with estimations on divergence events and demographic trends. *Study II* provides the first draft reference genome on the roan antelope and its use for intraspecific analyses at the genomic level. **Chapter 4** evaluates population genomic patterns across the roan and the sable antelope, through a comparative approach.

Chapter 5 regards a *General Discussion* on the several scientific findings throughout the thesis, challenges, and prospects.

1.6 References

Abdelkrim, J., Robertson, B. C., Stanton, J. A. L. and Gemmell, N. J., 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46: 185–192.

Abebe, B., Acocella, V., Korme, T. and Ayalew, D., 2007. Quaternary faulting and volcanism in the main Ethiopian Rift. *Journal of African Earth Sciences*, 48: 115–124.

Allard, M. W., Miyamoto, M. M., Jarecki, L., Kraus, F. and Tennant, M. R., 1992. DNA systematics and evolution of the artiodactyl family Bovidae. *Proceedings of the National Academy of Sciences*, 89: 3972–3976.

Allio, R., Donega, S., Galtier, N. and Nabholz, B., 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular. *Molecular Biology and Evolution*, 34: 2762–2772.

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Ansell, W. F. H., 1972. Order Artiodactyla, pp. 15–83 in *Mammals of Africa: an Identification Manual*. Smithsonian Institute Press (J. Meester and H. W. Setzer, Eds.), Washington, DC.

Arbogast, B. S. and Kenagy, G. J., 2001. Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, 28: 819-825.

Arctander, P., Johansen, C. and Coutellec-Vreto, M.-A., 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution*, 16: 1724–1739.

Arimoto, R., 2001. Eolian dust and climate: relationships to sources, tropospheric chemistry, transport and deposition. *Earth-Science Reviews*, 54: 29–42.

Armitage, S. J., Bristow, C. S. and Drake, N. A., 2015. West African monsoon dynamics inferred from abrupt fluctuations of Lake Mega-Chad. *Proceedings of the National Academy of Sciences*, 112: 8543–8548.

Asner, G. P., Elmore, A. J., Olander, L. P., Martin, R. E. and Harris, T., 2004. Grazing systems, ecosystem responses, and global change. *Annual Review of Environment and Resources*, 29: 261–299.

Asner, G. P., Levick, S., Kennedy-Bowdoin, T., Knapp, D., Emerson, R., Jacobson, J., Colgan, M. S. and Martin, R., 2009. Large-scale impacts of herbivores on the structural diversity of African savannas. *Proceedings of the National Academy of*

Sciences, 106: 4947–4952.

Avise, J., Arnold, J., Ball, R. and Bermingham, E., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18: 489–522.

Avise, J. C., 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, 7: 371–379.

Avise, J. C., Walker, D. and Johns, G. C., 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 265: 1707–1712.

Avise, J. C., 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36: 3–15.

Badgley, C., 2010. Tectonics, topography, and mammalian diversity. *Ecography*, 33: 220–231.

Ballard, J. W. O. and Whitlock, M. C., 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, 13: 729-744.

Beckage, B., Platt, W. J. and Gross, L. J., 2009. Vegetation, fire, and feedbacks: a disturbance-mediated model of savannas. *The American Naturalist*, 174: 805–818.

Beheregaray, L. B., 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17: 3754–3774.

Bermingham, E. and Moritz, C., 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology*, 7: 367–369.

Beuning, K., Kelts, K., Russell, J. and Wolfe, B., 2002. Reassessment of Lake Victoria–Upper Nile River paleohydrology from oxygen isotope records of lake-sediment cellulose. *Geology*, 30: 559–562Bibi, F., 2009. The fossil record and evolution of Bovidae: the state of the field. *Palaeontologia Electronica*, 12: 1–11.

Bibi, F., 2011. Mio-Pliocene faunal exchanges and African biogeography: the record of fossil bovids. *Plos One*, 6: e16688.

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

Bibi, F. and Kiessling, W., 2015. Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences*, 112: 10623–10628.

Birungi, J. and Arctander, P., 2000. Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, kob (*Kobus kob*). *Molecular Ecology*, 9: 1997–2008.

Bleidorn, C., 2016. Third generation sequencing: technology and its potential impact on

evolutionary biodiversity research. *Systematics and Biodiversity*, 14: 1–8.

Bobe, R. and Eck, G., 2001. Responses of African bovids to Pliocene climate change. *Paleobiology*, 27: 1–47.

Bobe, R and Behrensmeyer, A., 2004. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. *Paleoecology*, 207: 399–420.

Botero-Castro, F., Tilak, M., Justy, F., Catzeflis, F., Delsuc, F. and Douzery, E., 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats. *Molecular Phylogenetics and Evolution*, 69: 728–739.

Bothma, J. du P. and van Rooyen, N., 2005. *Intensive wildlife production in southern Africa*. Van Schaik Publishers (J. du P. Bothma and N. van Rooyen, Eds.), Pretoria.

Brashares, J., Garland, T. J. and Arcese, P., 2000. Phylogenetic analysis of coadaptation in behaviour, diet, and body size in the African antelope. *Behavioural Ecology*, 11: 452–463.

Bro-Jørgensen, J., 2011. Intra-and intersexual conflicts and cooperation in the evolution of mating strategies: lessons learnt from ungulates. *Evolutionary Biology*, 38: 28–41.

Bro-Jørgensen, J., 2016. Our Antelope Heritage - Why the Fuss?, pp. 1–10 in *Antelope Conservation: from Diagnosis to Action*. John Wiley & Sons, Ltd. (J. Bro-Jørgensen and P. M. David, Eds.).

Broughton, R. E., Milam, J. E. and Roe, B. A., 2001. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Research*, 11: 1958–1967.

Brown, D. M., Brenneman, R. A., Koepfli, K.-P., Pollinger, J. P., Milá, B., Georgiadis, N. J., Louis, E. E. Grether, G. F., Jacobs, D. K. and Wayne, R. K., 2007. Extensive population genetic structure in the giraffe. *BMC Biology*, 5:57–70.

Bruford, M. W. and Wayne, R. K., 1993. Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development*, 3: 939–943.

Brumfield, R., Beerli, P., Nickerson, D. and Edwards, S., 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, 18: 249–256.

Burrough, S. and Thomas, D., 2008. Late Quaternary lake-level fluctuations in the Mababe Depression: Middle Kalahari palaeolakes and the role of Zambezi inflows. *Quaternary Research*, 69: 388–403.

Burrough, S., Thomas, D. and Bailey, R., 2009. Mega-Lake in the Kalahari: a Late Pleistocene record of the Palaeolake Makgadikgadi system. *Quaternary Science*

Reviews, 28: 1392–1411.

Chardonnet, P. and Crosmary W., 2013. *Hippotragus equinus* Roan Antelope, pp. 548–556 in *Mammals of Africa VI*. Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Charlesworth, B., Sniegowski, P. and Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371: 215–220.

Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., et al., 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, 364: eaav6202.

Chikhi, L. and Bruford, M., 2005. Mammalian population genetics and genomics, pp. 539–584 in *Mammalian Genomics.* CABI (A. Ruvinsky and J. Marshall Graves, Eds.), Cambridge.

Chorowicz, J., 2005. The East African rift system. *Journal of African Earth Sciences*, 43: 379–410.

Codron, D., Codron, J., Lee-Thorp, J. A., Sponheimer, M., De Ruiter, D., Sealy, J., Grant, R. and Fourie, N., 2007. Diets of savanna ungulates from stable carbon isotope composition of faeces. *Journal of Zoology*, 273: 21–29.

Cohen, A. S., Stone, J. R., Beuning, K. R. M., Park, L. E., Reinthal, P. N., Dettman, D., Scholz, C. A., Johnson, T. C., et al., 2007. Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proceedings of the National Academy of Sciences*, 104: 16422–16427.

Cui, P., Ji, R., Ding, F., Qi, D., Gao, H., Meng, H., Yu, J., Hu, S. and Zhang, H., 2007. A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): an evolutionary history of Camelidae. *BMC Genomics*, 8: 241–251.

Curole, J. and Kocher, T., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution*, 14: 394–398.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques*, 52: 87–94.

Dabney, J, Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., et al., 2013. Complete mitochondrial genome sequence of a middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763.

Delvaux, D., Kervyn, F., Macheyeki, A. and Temu, E., 2012. Geodynamic significance of the TRM segment in the East African Rift (W-Tanzania): active tectonics and paleostress in the Ufipa plateau and Rukwa basin. *Journal of Structural Geology*,

37: 161–180.

DeMenocal, P. B. and Rind, D., 1993. Sensitivity of Asian and African climate to variations in seasonal insolation, glacial ice cover, sea surface temperature, and Asian orography. *Journal of Geophysical Research*, 98: 7265–7287.

DeMenocal, P. B., 1995. Plio-Pleistocene African climate. *Science*, 270: 53–59.

DeMenocal, P. B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, 220: 3–24.

DeMiguel, D., Azanza, B. and Morales, J., 2014. Key innovations in ruminant evolution: a paleontological perspective. *Integrative Zoology*, 9: 412-433.

de Wit, M., 1999. Post-Gondwana drainage and the development of diamond placers in western South Africa. *Economic Geology*, 94: 721–740.

Dörgeloh, W. G., 2001. A draft habitat suitability model for roan antelope in the Nylsvley Nature Reserve, South Africa. *African Journal of Ecology*, 39: 313–316.

Doucouré, C. and de Wit, M., 2003. Old inherited origin for the present near-bimodal topography of Africa. *Journal of African Earth Sciences*, 36: 371–388.

Dupont, L., Schmüser, A., Jahns, S. and Schneider, R., 1999. Marine-terrestrial interaction of climate changes in West Equatorial Africa of the last 190,000 years. *Palaeoecology of Africa*, 26: 61–84.

Dupont, L, and Wyputta, U., 2003. Reconstructing pathways of aeolian pollen transport to the marine sediments along the coastline of SW Africa. *Quaternary Science Reviews*, 22: 157–174.

Dupont, L., 2011. Orbital scale vegetation change in Africa. *Quaternary Science Reviews*, 30: 3589–3602.

Du Toit, J. T. and Cumming, D. H. M., 1999. Functional significance of ungulate diversity in African savannas and the ecological implications of the spread of pastoralism. *Biodiversity and Conservation*, 8: 1643–1661.

East, R. and IUCN SSC Antelope Specialist Group, 1999. *African antelope database 1998* (IUCN/SSC Antelope Specialist Group, Ed.), Gland: IUCN.

Eblate, E. M., Lughano, K. J., Sebastian, C. D., Peter, M. L. and Knut, R. H., 2011. Polymorphic microsatellite markers for genetic studies of African antelope species. *African Journal of Biotechnology*, 10: 11817–11820.

Edmands, S., 2006. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, 16: 463–475.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5: 435–445.

Espregueira Themudo, G., Rufino, A. C. and Campos, P. F., 2015. Complete

mitochondrial DNA sequence of the endangered giant sable antelope (*Hippotragus niger variani*): Insights into conservation and taxonomy. *Molecular Phylogenetics and Evolution*, 8:, 242–249.

Estes, R., 1991. The significance of horns and other male secondary sexual characters in female bovids. *Applied Animal Behaviour Science*, 29: 403–451.

Estes, R., 2013. *Hippotragus niger* Sable antelope, pp. 556–565 in *Mammals of Africa VI.* Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Estoup, A., Jarne, P. and Cornuet, J. M., 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11: 1591-1604.

Etherington, G., Heavens, D., Baker, D., Lister, A., McNelly, R., Garcia, G., Clavijo, B., Macaulay, I., Haerty, W. and di Palma, F., 2020. Sequencing smart: de novo sequencing and assembly approaches for a non-model mammal. *GigaScience*, 9: giaa045.

Faith, J. T. and Behrensmeyer, A., 2013. Climate change and faunal turnover: testing the mechanics of the turnover-pulse hypothesis with South African fossil data. *Paleobiology*, 39: 609–627.

Faith, J. T. and Thompson, J. C., 2013. Fossil evidence for seasonal calving and migration of extinct blue antelope (*Hippotragus leucophaeus*) in southern Africa. *Journal of Biogeography*, 40: 2108–2118.

Fernández, H. and Vrba, E., 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews*, 80: 269–302.

Funk, W. C., McKay, J. K., Hohenlohe, P. A. and Allendorf, F. W., 2012. Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27: 489–496.

Galtier, N., Nabholz, B., Glémin, S. and Hurst, G. D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18: 4541-4550.

Gardner, M. G., Fitch, A. J., Bertozzi, T. and Lowe, A. J., 2011. Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, 11: 1093–1101.

Goudie, A. S., 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67: 437–456.

Grobler, J. and Nel, G., 1996. Lack of allozyme heterogeneity at eight loci studied in a roan antelope *Hippotragus equinus* population from the Percy Fyfe Nature Reserve. *South African Journal of Wildlife*, 26: 34–35.

Grover, A. and Sharma, P., 2016. Development and use of molecular markers: past and present. *Critical Reviews in Biotechnology*, 36: 290–302.

Guilderson, T., Fairbanks, R. and Rubenstone, J., 1994. Tropical temperature variations since 20 000 years ago: modulating interhemispheric climate change. *Science*, 263: 663–665.

Hagelberg, E., 2003. Recombination or mutation rate heterogeneity? Implications for the Mitochondrial Eve. *Trends in Genetics*, 19: 84–90.

Harrington, R., Owen-Smith, N., Viljoen, P. C., Biggs, H. C., Mason, D. R. and Funston, P., 1999. Establishing the causes of the roan antelope decline in the Kruger National Park, South Africa. *Biological Conservation*, 90: 69–78.

Hassanin, A. and Douzery, E., 1999. The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial cytochrome b gene. *Molecular Phylogenetics and Evolution*, 13: 227–243.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., et al., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies*, 335: 32-50.

Hewitt, G., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58: 247–276.

Hewitt, G., 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68: 87–112.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405: 907–913.

Hewitt, G., 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, 10: 537–549.

Hewitt, G., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359: 183–195.

Hewitt, G., 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139: 617–638.

Hiendleder, S., Lewalski, H. and Janke, A., 2008. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenetic and Genome Research*, 120: 150–156.

Hodgkinson, A. and Eyre-Walker, A., 2011. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics*, 12: 756–766.

Hooghiemstra, H., Lézine, A., Leroy, S., Dupont, L. and Marret, F., 2006. Late Quaternary palynology in marine sediments: a synthesis of the understanding of pollen distribution patterns in the NW African setting. *Quaternary International*, 148:

29–44.

Hume, I., 2002. Digestive strategies of mammals. *Acta Zoologica Sinica*, 48: 1–19.

IUCN SSC Antelope Specialist Group, 2008. *Hippotragus niger* ssp. *variani*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10169A50188611.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017a. *Hippotragus equinus*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10167A50188287.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017b. *Hippotragus niger*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10170A50188654.en on April 8th, 2020,

Janecek, L., Honeycutt, R., Adkins, R. and Davis, S., 1996. Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily Bovinae. *Molecular Phylogenetics and Evolution*, 6: 107–119.

Jansen, E. and Sjøholm, J., 1991. Reconstruction of glaciation over the past 6 Myr from ice-borne deposits in the Norwegian Sea. *Nature*, 349: 600–603.

Jansen van Vuuren, B. J. and Robinson, T. J., 2001. Retrieval of four adaptive lineages in duiker antelope: evidence from mitochondrial DNA sequences and fluorescence *in situ* hybridization. *Molecular Phylogenetics and Evolution*, 20: 409–425.

Jansen van Vuuren, B., Robinson, T. J., Vaz Pinto, P., Estes, R. and Matthee, C. A., 2010. Western Zambian sable: are they a geographic extension of the giant sable antelope? *South African Journal of Wildlife Research*, 40: 35–42.

Keeley, J. E. and Rundel, P. W., 2005. Fire and the Miocene expansion of C4 grasslands. *Ecology Letters*, 8: 683–690.

Kerley, G. I. H., Sims-Castley, R., Boshoff, A. F. and Cowling, R. M., 2009. Extinction of the blue antelope *Hippotragus leucophaeus*: modelling predicts non-viable global population size as the primary driver. *Biodiversity and Conservation*, 18: 3235–3242.

Kholodova, M. V, Lushchekina, A. A., Severtsov, A. N., Easton, A. J., Amgalan, L., Arylov, I. A., Bekenov, A., Grachev, I. A, Ryder, O. and Milner-Gulland, E. J., 2006. Mitochondrial DNA variation and population structure of the critically endangered saiga antelope *Saiga tatarica*. *Oryx*, 40: 103–107.

Kimanzi, J. K. and Wanyingi, J. N., 2014. The declining endangered roan antelope population in Kenya: what Is the way forward? In *Conference Papers in Science,* Hindawi.

King, S. D. and Ritsema, J., 2000. African hot spot volcanism: small-scale convection in the upper mantle beneath cratons. *Science*, 290: 1137–1140.

Kingston, J. and Harrison, T., 2007. Isotopic dietary reconstructions of Pliocene herbivores at Laetoli: implications for early hominin paleoecology. *Paleoecology*, 243: 272–306.

Klein, R. G., 1974. The taxonomic status, distribution and ecology of the blue antelope, *Hippotragus leucophaeus* (Pallas, 1766). *Annals of the South African Museum*, 65: 99–143.

Knapp, M. and Hofreiter, M., 2010. Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes*, 1: 227–243.

Kröger, R. and Rogers, K. H., 2005. Roan (*Hippotragus equinus*) population decline in Kruger National Park, South Africa: influence of a wetland boundary. *European Journal of Wildlife Research*, 51: 25–30.

Kuznetsova, M. V., Kholodova, M. V. and Luschekina, A. A., 2002. Phylogenetic analysis of sequences of the 12S and 16S rRNA mitochondrial genes in the family Bovidae: new evidence. *Russian Journal of Genetics*, 38: 942–950.

Kyalangalilwa, B., Boatwright, J. S., Daru, B. H., Maurin, O. and van der Bank, M., 2013. Phylogenetic position and revised classification of *Acacia* s.l. (Fabaceae: Mimosoideae) in Africa, including new combinations in *Vachellia* and *Senegalia*. *Botanical Journal of the Linnean Society*, 172: 500-523.

Lamb, H., Bates, C., Coombes, P., Marshall, M., Umer, M., Davies, S. and Dejen, E., 2007. Late Pleistocene desiccation of Lake Tana, source of the Blue Nile. *Quaternary Science Reviews*, 26: 287–299.

Lander, E., 1996. The new genomics: global views of biology. *Science*, 274: 536–539.

Lézine, A., Hély, C., Grenier, C., Braconnot, P. and Krinner, G., 2011. Sahara and Sahel vulnerability to climate changes, lessons from Holocene hydrological data. *Quaternary Science Reviews*, 30: 3001–3012.

Li, Y., Korol, A. B., Fahima, T., Beiles, A. and Nevo, E., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11: 2453-2465.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J. and Rahbek, C., 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39: 1189–1205.

Lindsey, P. A., Barnes, J., Nyirenda, V., Pumfrett, B., Tambling, C. J., Taylor, W. A. and Rolfes, M. T. S., 2013. The Zambian wildlife ranching industry: scale, associated benefits, and limitations affecting its development. *Plos One*, 8: e81761.

Lister, A. M., 2004. The impact of Quaternary Ice Ages on mammalian evolution. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359: 221–241.

Liu, N., Chen, L., Wang, S., Oh, C. and Zhao, H., 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6: S26.

Livingston, G. and Kingdon, J., 2013. The evolution of a continent: geography and geology, pp. 27–42 in *Mammals of Africa I.* A&C Black (K. J, D. Happold, M. Hoffmann, T. Butynski, M. Happold and J. Kalina, Eds.).

Loison, A., Gaillard, J., Pélabon, C. and Yoccoz, N., 1999. What factors shape sexual size dimorphism in ungulates? *Evolutionary Ecology Research*, 1: 611–633.

Lorenzen, E. D., Heller, R. and Siegismund, H. R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21: 3656–3670.

Luikart, G. and England, P. R., 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology & Evolution*, 14: 253–256.

Luikart, Gordon, England, P. R., Tallmon, D., Jordan, S. and Taberlet, P., 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4: 981–994.

Lyons, R. P., Scholz, C. A., Cohen, A. S., King, J. W., Brown, E. T., Ivory, S. J., Johnson, T. C., Deino, A. L., et al., 2015. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proceedings of the National Academy of Sciences*, 112: 15568–15573.

Mamanova, L., Coffey, A., Scott, C., Kozarewa, I., Turner, E., Kumar, A., Howard, E., Shendure, J. and Turner, D., 2010. Target enrichment strategies for next generation sequencing. *Nature Methods*, 7: 111–118.

Maricic, T., Whitten, M. and Pääbo, S., 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *Plos One*, 5: e14004.

Matthee, C. A. and Robinson, T. J., 1999. Mitochondrial DNA population structure of roan and sable antelope; implications for the translocation and conservation of the species. *Molecular Ecology*, 8: 227–238.

Matthee, C. A. and Davis, S. K., 2001. Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective. *Molecular Biology and Evolution*, 18: 1220–1230.

Matthee, C. A., Burzlaff, J. D., Taylor, J. F. and Davis, S. K., 2001. Mining the mammalian genome for Artiodactyl systematics. *Systematic Biology*, 50: 367–390.

Maxwell, T. A. and Haynes, C. V., 2010. Evidence for Pleistocene lakes in the Tushka region, south Egypt. *Geology*, 38: 1135–1138.

Mayaux, P., Bartholomé, E., Fritz, S. and Belward, A., 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31: 861–877.

McCarthy, T. S. and Ellery, W. N., 1998. The Okavango delta. *Transactions of the Royal*

Society of South Africa, 53: 157–182.

Mendoza, M. and Palmqvist, P., 2008. Hypsodonty in ungulates: an adaptation for grass consumption or for foraging in open habitat? *Journal of Zoology*, 274: 134–142.

Mittal, N. and Dubey, A., 2009. Microsatellite markers – a new practice of DNA based markers in molecular genetics. *Pharmacognosy Reviews*, 3: 235–246.

Moore, A. E. and Larkin, P., 2001. Drainage evolution in south-central Africa since the breakup of Gondwana. *South African Journal of Geology*, 104: 47–68.

Moore, A. T. and Blenkinsop, T., 2002. The role of mantle plumes in the development of continental-scale drainage patterns: the southern African example revisited. *South African Journal of Geology*, 105: 353–360.

Moore, A. E., Cotterill, F. and Eckardt, F., 2012. The evolution and ages of Makgadikgadi palaeo-lakes: consilient evidence from Kalahari drainage evolution south-central Africa. *South African Journal of Geology*, 115: 385–413.

Moorley, R. and Kingdon, J., 2013. Africa's environmental and climatic past, pp. 43–56 in *Mammals of Africa I.* Bloomsbury Publishing (K. J, D. Happold, M. Hoffmann, T. Butynski, M. Happold and J. Kalina, Eds.), London.

Morin, P. A., Chambers, K. E., Boesch, C. and Vigilant, L., 2001. Quantitative polymerase chain reaction analysis of DNA from non-invasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, 10: 1835–1844.

Moritz, C., 1994. Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, 9: 373–375.

Moritz, C., 1999. Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas*, 130: 217-228.

Nachman, M., 2002. Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics & Development*, 12: 657–663.

Nersting, L. G. and Arctander, P., 2001. Phylogeography and conservation of impala and greater kudu. *Molecular Ecology*, 10: 711–719.

Nokelainen, O. and Stevens, M., 2016. Camouflage. *Current Biology*, 26: R654–R656.

Nyakaana, S., Arctander, P. and Siegismund, H. R., 2002. Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity*, 89: 90–98.

Pachur, H. and Rottinger, F., 1997. Evidence for a large extended paleolake in the Eastern Sahara as revealed by spaceborne radar lab images. *Remote Sensing of Environment*, 61: 437–440.

Parrini, F. and Owen-Smith, N., 2010. The importance of post-fire regrowth for sable antelope in a Southern African savanna. *African Journal of Ecology*, 48: 526–534.

Peel, M., Finlayson, B. and Mcmahon, T., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11: 1633–1644.

Pepin, L., Amiguest, Y., Lepingle, A., Berthier, J.-L., Bensaid, A. and Vaiman, D., 1995. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heredity*, 74: 53–61.

Pérez-Barbería, F. J., Gordon, I. J. and Illius, A., 2001. Phylogenetic analysis of stomach adaptation in digestive strategies in African ruminants. *Oecologia*, 129: 498–508.

Pérez-Barbería, F. J., Gordon, I. J. and Pagel, M., 2002. The origins of sexual dimorphism in body size in ungulates. *Evolution*, 56: 1276–1285.

Pik, R., Marty, B. and Hilton, D., 2006. How many mantle plumes in Africa? The geochemical point of view. *Chemical Geology*, 226: 100–114.

Piltz, J., Sorensen, T. and Ferrie, G. M., 2016. Population analysis and breeding and transfer plan: sable antelope (*Hippotragus niger*). AZA Species Survival Plan Yellow Program. Report by *AZA Population Management Center*.

Pitra, C., Hansen, A. J., Lieckfeldt, D. and Arctander, P., 2002. An exceptional case of historical outbreeding in African sable antelope populations. *Molecular Ecology*, 11: 1197–1208.

Pitra, C., Vaz Pinto, P., O'Keeffe, B. W. J., Willows-Munro, S., Jansen van Vuuren, B. and Robinson, T. J., 2006. DNA-led rediscovery of the giant sable antelope in Angola. *European Journal of Wildlife Research*, 52: 145–152.

Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E. and Gill, T. E., 2002. Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. *Reviews of Geophysics*, 40: 212–231.

Prothero, D., 1993. Ungulate phylogeny: molecular vs. morphological evidence. *Mammal Phylogeny*, 2: 173–181.

Raymo, M. E., 1994. The initiation of Northern Hemisphere glaciation. *Annual Review of Earth & Planetary Sciences*, 22: 353–383.

Rebholz, W. and Harley, E., 1999. Phylogenetic relationships in the bovid subfamily Antilopinae based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 12: 87–94.

Richard, G.-F., Kerrest, A. and Dujon, B., 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews*, 72: 686–727.

Rienzo, A. Di, Peterson, A. C., Garzat, J. C., Valdes, A. M., Slatkint, M. and Freimer, N.

B., 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, 91: 3166–3170.

Roberts, E. M., Stevens, N. J., O'connor, P. M., Dirks, P. H. G. M., Armstrong, W. C. and Kemp, R. A., 2012. Initiation of the western branch of the East African Rift coeval with the eastern branch. *Nature Geoscience*, 5: 289–294.

Rocha, J., 2014. The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences. MsC thesis dissertation, University of Porto, Portugal.

Rokas, A., Ladoukakis, E. and Zouros, E., 2003. Animal mitochondrial DNA recombination revisited. *Trends in Ecology & Evolution*, 18: 411–417.

Rokas, A. and Carroll, S., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*, 22: 1337–1344.

Rubinsztein, D., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.-H., Margolis, R. L., Ross, C. A. and Ferguson-Smith, M., 1995. Microsatellite evolution – evidence for directionality and variation in rate between species. *Nature Genetics*, 10: 337–343.

Schlötterer, C. and Tautz, D., 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20: 211–215.

Schlötterer, C., 2004. The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics*, 5: 63–69.

Scholz, C. A., Johnson, T. C., Cohen, A. S., King, J. W., Peck, J. A., Overpeck, J. T., Talbot, M. R., Brown, E. T., et al., 2007. East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proceedings of the National Academy of Sciences*, 104: 16416–16421.

Sclater, P. and Thomas, O., 1899. The Blue-Buck, pp. 4–12 in *The Book of Antelopes IV*. Taylor & Francis (P. Sclater and O. Thomas, Eds.), London.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., et al., 1997. Modelling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275: 502–509.

Shastry, B. S., 2002. SNP alleles in human disease and evolution. *Journal of Human Genetics*, 47: 0561–0566.

Shaw, A. I. and Goudie, A. S., 2002. Geomorphological evidence for the extension of the Mega-Kalahari into south-central Angola. *South African Geographical Journal*, 84: 182–194.

Sigurðardóttir, S., Helgason, A., Gulcher, J., Stefansson, K. and Donnelly, P., 2000. The mutation rate in the human mtDNA control region. *The American Journal of Human*

Genetics, 66: 1599–1609.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87: 651–701.

Stager, J. C. and Johnson, T. C., 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia*, 596: 5–16.

Stankiewicz, J. and de Wit, M., 2006. A proposed drainage evolution model for Central Africa – did the Congo flow east? *Journal of African Earth Sciences*, 44: 75–84.

Stankowich, T. and Caro, T., 2009. Evolution of weaponry in female bovids. *Proceedings of the Royal Society B: Biological Sciences*, 276: 4329–4334.

Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M. and Meyer, M., 2009. Direct multiplex sequencing (DMPS)-a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Research*, 19: 1843–1848.

Stoner, C., Caro, T. and Graham, C., 2003. Ecological and behavioural correlates of coloration in artiodactyls: systematic analyses of conventional hypotheses. *Behavioural Ecology*, 14: 823–840.

Sunnucks, P., 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, 15: 199–203.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. and Cosson, J. F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7: 453–464.

Timberlake, J. and Chidumayo, E., 2011. Miombo ecoregion vision report. *Occasional Publications in Biodiversity*, 20.

The International SNP Map Working Group, C. S. H. L., Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., et al., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409: 928–934.

Toews, D. P. and Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21: 3907-3930.

Trauth, M. H., Deino, A., Bergner, A. and Strecker, M., 2003. East African climate change and orbital forcing during the last 175 kyr BP. *Earth and Planetary Science Letters*, 206: 297–313.

Trauth, M. H., Deino, A., Center, B. G. and Strecker, M. R., 2005. Late Cenozoic moisture history of East Africa. *Science*, 309: 2051–2053.

Trauth, M. H., Maslin, M., Deino, A., Strecker, M., Bergner, A. and Dühnforth, M., 2007. High-and low-latitude forcing of Plio-Pleistocene East African climate and human

evolution. *Journal of Human Evolution*, 53: 475–486.

Trauth, M. H., Larrasoaña, J. C. and Mudelsee, M., 2009. Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, 28: 399–411.

Väli, Ü., Einarsson, A., Waits, L. and Ellegren, H., 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, 17: 3808–3817

van Hooft, W. F., Groen, A. F. and 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: 267–279.

van Wyk, A. M., Dalton, D. L., Kotzé, A., Grobler, P. J., Mokgokong, P. S., Kropff, A. S. and Jansen van Vuuren, B., 2019. Assessing introgressive hybridization in roan antelope (*Hippotragus equinus*): lessons from South Africa. *Plos One*, 14: e0213961.

Vaz Pinto, P., 2009. Giant sable rescue and translocation. *Gnusletter*, 28: 8–10.

Vaz Pinto, P., Lopes, S., Mourão, S., Baptista, S., Siegismund, H. R., Jansen van Vuuren, B., Beja, P., Ferrand, N. and Godinho, R., 2015. First estimates of genetic diversity for the highly endangered giant sable antelope using a set of 57 microsatellites. *European Journal of Wildlife Research*, 61: 313–317.

Vaz Pinto, P., Beja, P., Ferrand, N. and Godinho, R., 2016. Hybridization following population collapse in a critically endangered antelope. *Scientific Reports*, 6: 18788.

Vaz Pinto, P., 2018. Evolutionary history of the critically endangered giant sable antelope (*Hippotragus niger variani*). Insights into its phylogeography, population genetics. PhD thesis dissertation, University of Porto, Portugal.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., et al., 2001. The sequence of the human genome. *Science*, 291: 1304–1351.

Vignal, A., Milan, D., Sancristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34: 275–305.

Viguera, E., Canceill, D. and Ehrlich, S., 2001. Replication slippage involves DNA polymerase pausing and dissociation. *The EMBO Journal*, 20: 2587–2595.

Vrba, E., 1996. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate, pp. 385–424 in *Paleoclimate and Evolution*. Yale University Press (E. Vrba, G. Denton, T. Partidge and L. Burckle, Eds.), Connecticut.

Wan, Q.-H., Wu, H., Fujihara, T. and Fang, S.-G., 2004. Which genetic marker for which

conservation genetics issue? *Electrophoresis*, 25: 2165–2176.

Weckerly, F. W., 1998. Sexual-size dimorphism: influence of mass and mating systems in the most dimorphic mammals. *Journal of Mammalogy*, 79: 33–52.

White, F., 1983. *The vegetation of Africa*. United Nations Educational, Scientific and Cultural Organization (Unesco, Ed.), Paris.

Wright, J. D., 2000. Global climate change in Marine Stable Isotope records, pp. 671–682 in *Quaternary Geochronology: Methods and Applications*. American Geophysical Union (J. S. Noller, J. M. Sowers and W. R. Lettis, Eds.).

Xia, P., Radpour, R., Zachariah, R. and Fan, A., 2009. Simultaneous quantitative assessment of circulating cell-free mitochondrial and nuclear DNA by multiplex real-time PCR. *Genetics and Molecular Biology*, 32: 20–24.

Yang, Z. and Nielsen, R., 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution*, 46: 409-418.

Yu, L., Li, Y. W., Ryder, O. A. and Zhang, Y. P., 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evolutionary Biology*, 7: 198–209.

Zane, L., Bargelloni, L. and Patarnello, T., 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11: 1-16.

Zhang, D. X. and Hewitt, G. M., 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, 12: 563-584.

Zhivotovsky, L. A. and Feldmant, M. W., 1995. Microsatellite variability and genetic distances. *Proceedings of the National Academy of Sciences*, 92: 11549–11552.

CHAPTER 2

Review on genetic discontinuities across Sub-Saharan Africa

Genetic discontinuities of fauna and flora across African savannas: a review

This work was developed within the ECOGEN research group[1], with the participation of every members, namely Mariana Ribeiro, Rita Rocha, Pedro Silva, Joana Rocha, Diana Lobo, Carolina Pacheco, Maris Hindrikson, Gonçalo Ferraz, and Margarida Andrade, with conception and main supervision by Nuno Ferrand and Raquel Godinho, as well as valuable input from Pedro Tarroso[1] and Laura Tensen[2].

My contribution involved the extensive literature search, data analysis, figure illustrations, and writing. Similar contributions were made from Mariana Ribeiro and Rita Rocha.

Affiliation's: [1]CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus de Vairão, Portugal; [2]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South-Africa

2.1 Introduction

Sub-Saharan Africa is one of the biologically richest areas in the world. Two main biomes dominate this region: tropical forests and savannas. The tropical forest belt, located across Central Africa, in the Congo basin, is the second largest area of contiguous moist tropical forest in the World, covering about 1.8 million km^2 (Mayaux et al. 2013). The region is characterized by year-round high precipitation, leading to a dense tree cover. It shelters some of the greatest biological diversity of African fauna and flora, playing a vital role in the global carbon-cycle (Bowman 2000; Grace et al. 2006) and worldwide ecological services (de Wasseige et al. 2014). In coastal West Africa and parts of East Africa several patches of tropical forest also dominate the landscape (Mayaux et al. 2013). Savannas are defined by the coexistence of a continuous grass layer with a discontinuous tree cover (Ratnam et al. 2011; Parr et al. 2014), encompassing grasslands, scrublands, and woodlands, driven by low rainfall and pronounced dry seasons (Jacobs 2004). Savanna areas are more widespread than tropical forests, covering about 50% of the African continent, through more than 15 million km^2 (Grace et al. 2006). They constitute one of the World's major biomes (Kim and Eltahir 2004), harbouring the largest remaining mammalian megafauna on Earth (Fjeldså et al. 2004). The distribution and structure of both forest and savanna biomes results from the

interaction between climate, soil, and disturbance regimes, namely fire and herbivory (Sankaran et al. 2005).

Understanding the mechanisms that underlie patterns of biodiversity for a given biome is a challenging task. For genetic diversity, patterns across geographic areas are studied in the scope of the Phylogeography discipline (Avise 2000, 2009; Kidd and Ritchie 2006; Hickerson et al. 2010). Congruent phylogeographical patterns across multiple co-distributed lineages may suggest similar mechanisms of diversification and general evolutionary drivers for the region (Bermingham and Moritz 1998; Hickerson, Dolman and Moritz 2006; Arbogast and Kenagy 2008). This has been studied in different comparative studies across several regions of the World, which have unravelled common patterns of diversity distribution, for example in Europe (Taberlet et al. 1998; Hewitt 2000; Schmitt 2007), North-America (Soltis et al. 2006; Shafer et al. 2010), Central America (Bagley and Johnson 2014), South-America (Leite and Rogers 2013), and Australia (Rix et al. 2015). These and many more studies have identified the climatic oscillations of the Pleistocene period (~11 kya to 2.5 Mya) as one of the most important processes shaping current distribution of genetic diversity. Alternating climatic cycles promoted the repeated contraction and expansion of populations, isolating and acting as refugia or promoting admixture, respectively (Hewitt 1996, 1999; Waltari et al. 2007). Across sub-Saharan Africa, patterns of biodiversity for both tropical forests (Plana 2004; Hardy et al. 2013; Couvreur et al. 2020) and savanna biomes (DeMenocal 2004; Hewitt 2004; Lorenzen, Heller and Siegismund 2012) have been highly correlated with historical climatic and geomorphological changes. Due to a drastic continental cooling and aridification during late-Pliocene (1.5-3.5 Mya) (Morley 2011; Kissling et al. 2012; Linder 2017), Africa witnessed a gradual change of vegetation cover, with the replacement of large extensions of tropical forests by open woodlands and savannas (Bobe 2006; Dupont 2011; Kaya et al. 2018). This period was marked by the rapid evolution of savanna-adapted taxa, an emblematic example of which are the bovid species we found today (Vrba 1996; Bobe and Eck 2001; Bibi and Kiessling 2015). During the Pleistocene climatic cycles, however, African climates became increasingly variable (DeMenocal 1995; Trauth, Larrasoaña and Mudelsee 2009), with cyclic variations between cold and dry interpluvials, and warm and wet pluvial periods (Dupont 2011). During interpluvials, climatic conditions would favour the expansion of savannas over tropical forests, and during pluvials the opposite would occur (DeMenocal 2004; Feakins, DeMenocal and Eglinton 2005; Dupont 2011).

Africa has also had a dynamic geological history, with drastic landmass configurational changes in relatively recent times, during Plio- and Pleistocene periods. These include uplifting forces, rifting and volcanism in tectonically active regions of

Eastern and Southcentral Africa, that violently and repeatedly modified landscapes, fragmenting and reconnecting habitats for fauna and flora (Grove 1986; Doucouré and de Wit 2003; Moore, Blenkinsop and Cotterill 2009; Badgley 2010). One of the most striking geological features of the African continent is the East African Rift System (EARS), comprising a series of successive aligned, individual tectonic basins, extending for several thousands of kilometres through East and Southern Africa (Chorowicz 2005). Tectonic basins date back ca. 30 Myr, however much of the modern topography was generated throughout the last 2 Myr (Chorowicz 2005; Roberts et al. 2012; Jess et al. 2020). The EARS is composed by two main branches: the western branch, also called Albertine Rift, including the Great Lakes, such as Albert, Tanganyika, Malawi and Kiwu; and the eastern branch, also known as Gregory Rift, running along the Afar triangle, through the Ethiopian and the Kenyan Rifts, and ending in the basins across Tanzania, which integrate the Eastern Arc Mountains (EAM). The Rift System has a number of well-known volcanic peaks that are high in elevation, with permanent glaciers, despite close proximity to the Equator (Osmaston and Harrison 2005). The biggest continuous set of highlands constitute the Ethiopian Highlands, situated on the Ethiopian plateau (Aldenderfer 2006). On the western side of the continent, Mount Cameroon, in Central Africa, is formed by the volcanic activity of the Cameroon Volcanic Line (CVL). Such continental changes have also been driving the evolution of African drainage systems (Goudie 2005; Stankiewicz and de Wit 2006; Moore et al. 2009), defining the four largest hydrographic basins: Chad, Congo, Niger, and Nile (Thieme et al. 2005; Rebelo, McCartney and Finlayson 2010).

Comparative phylogeographic studies in sub-Saharan Africa have been limited to large-bodied mammals, namely bovids. For those taxa, spatially congruent phylogenetic patterns have allowed to establish three major clades placed in the West, East and Southern Africa (Hewitt 2004; Lorenzen et al. 2012). These regions are suggested to embrace each main glacial refugia during late-Pliocene and through the Pleistocene, sustaining suitable habitat and maintaining genetic diversity. The diversity of savanna ecosystems and evolutionary history of savanna-adapted taxa was, therefore, driven by speciation within, and dispersal events from such refugia (Hewitt 2004; Lorenzen et al. 2012). Each of these regions has experienced divergent environmental, climatic, and tectonic conditions, which is mirrored by the mosaic of landscapes, habitats, and taxa occurring at present times. The predominate evolutionary divergence occurs between West and Southern clades, which would become isolated by an extended tropical forest belt during pluvials (Bobe and Eck 2001; Bobe, Behrensmeyer and Chapman 2002; Smith 2012). For the Southern region, both palynological data (Scott, Steenkamp and Beaumont 1995) and less pronounced phylogeographic structuring, with higher levels of

genetic diversity across taxa, indicate the relative environmental and vegetational stability from the Pleistocene onwards (DeMenocal 2004; Trauth et al. 2005; Maslin et al. 2012), supporting its status as a long-standing savanna refugium. East Africa is described has representing the connection of long-diverged lineages during interpluvial range expansions (Hewitt 2004; Lorenzen et al. 2012). Climatic heterogeneity and complex topography, associated with the EARS, triggered evolutionary changes across East African populations, with natural selection and genetic drift acting together on isolated populations (DeMenocal 2004; Maslin and Christensen 2007; Badgley 2010; Lorenzen et al. 2012). There is also evidence of substructure within each main region, and even the suggestion of a fourth region across Southwest Africa (Lorenzen et al. 2012; Bertola et al. 2016).

Comparative analyses solely based on large-bodied mammals may scarcely reflect the differential ecological constraints experienced by the immense biological diversity beyond this taxon. Therefore, direct extrapolations of previous studies to other vertebrates, invertebrates or even plant taxa may lead to biased inferences (Avise 2000, 2009). Also, savanna biomes encompass different landscapes, namely grasslands, semi-arid savannas, non-arid woodlands and deciduous forests, and its division into three main regions may, therefore, be an oversimplification. For instance, paleo-environmental reconstructions suggest that the savanna corridor currently separating West from Central African rainforest blocks – known as the Dahomey Gap – was forested during pluvial periods, while open vegetation developed under drier conditions (Booth 1958; Salzmann and Hoelzmann 2005). This switching corridor has been pointed out as the major feature driving the phylogeographic structure of West/Central African species (Demenou, Piñeiro and Hardy 2016; Demenou, Doucet and Hardy 2018). Similarly, other climatic or geomorphological changes could have been equally important drivers of species evolution in savanna biomes. For example, the EARS has both been recognized as a major biogeographic barrier for animal dispersal, but as a hotspot for endemism, due to its topographic and environmental heterogeneity (Denys, Chorowicz and Tiercelin 1986; Chorowicz 2005; Burgess et al. 2007; Goodier et al. 2011; Castiglia et al. 2012; Lorenzen et al. 2012; Colangelo et al. 2013; Menegon et al. 2014). Drainage evolution across sub-Saharan Africa has also been pointed out as a main factor driving vicariant events for several taxa (Stankiewicz and de Wit 2006; McDonough et al. 2015; Petružela et al. 2018), but an overall evaluation of available data is lacking.

In this work, we review phylogenetic and phylogeographic studies in savanna species across their sub-Saharan distribution, where data have been accumulated but an overview of overall, as well as taxa-specific patterns, is missing. Specifically, we are aiming to address the following questions: i) are the three sub-Saharan regions, namely

West, East and Southern Africa, defined in previous reviews as the most relevant phylogeographic splits for large-bodied mammals, congruent across vertebrate (mammals, birds, reptiles, amphibians and fish), invertebrates and plant species? ii) which are the main climatic and geomorphological drivers of genetic diversity patterns within each of these three main regions across vertebrates and plants? iii) is there any east-west or north-south gradient in phylogeographic patterns? Additionally, we also aim to assess trends in the availability of data for the different taxa, and to acknowledge how data deficiency limits the evaluation of global phylogeographic patterns.

2.2 Material and Methods

2.2.1 Study-area

Our study area encompassed the savanna biome in continental Africa, excluding islands and Madagascar, as well as the Congo subregion within the Congolian biogeographic region, as defined by Linder et al. (2012), which corresponds to the core rainforest belt. We only considered sampling across this latter region if the taxa distribution area also encompassed at least one out of the three main savanna regions – West, East, and Southern (as defined by Hewitt 2004 and Lorenzen et al. 2012). Rainforest was not included in our study because previous studies have assessed main patterns of phylogeographic discontinuities within this region (Hardy et al. 2013; Couvreur et al. 2020).

2.2.2 Literature Review

We searched Web of Science database for articles assessing phylogenetic or phylogeographic patterns in our study area published until 31st December 2018 and written in English. We used multiple combinations of three keywords for site, subject and taxa. Site keyword was set to ‘Africa’ which was combined with seven subjects (‘population genetics’; ‘phylogeography’; ‘phylogeny’; ‘genetic patterns’; ‘genetic structure’; ‘evolutionary patterns’, ‘evolutionary history’ and ‘diversification patterns’), and nine taxa (‘mammals’; ‘birds’; ‘reptiles’; ‘amphibians’; ‘fish’; ‘insects’; ‘arthropods’, “plants” and ‘flora’). Only extant, resident, and non-invasive animals and plants from terrestrial or freshwater habitats were selected. Research on species that suffered recent and strong bottlenecks (e.g. the black rhinoceros, *Diceros bicornis*) was not considered. Our database was then complemented with a search at the literature of review articles. This search resulted in a compilation of 1,023 research articles. From this, only studies

sampling roughly more than 25% of the known current distribution of species under analysis were used, as well as those studies with information on sampling location, either provided by a legible map or a table. Research articles only focusing species whose distributional range is limited to a centre of endemism (Text S2.1), totalizing 99 articles (Table S2.1) were not considered towards evaluation of genetic discontinuities, since they would mainly recognize diversity at a micro-scale. Exceptions of this refer to discontinuities between a taxon with restricted distribution, and another with wider distribution.

Selected research articles were classified according to taxonomic group, sampling area and molecular data. Taxa were grouped into seven taxonomic groups, the same used as keywords, combining insects and arthropods into “invertebrates”, as well as plants and flora into “plants”. This latter group includes only angiosperm species, since no articles on gymnosperms passed through the criteria above. Our use of the term taxa includes genera, species, and species-complexes. Sampling area was recorded to countries, to sampled region, according to the study-area, and occurrence region, according to the currently known species distribution. Molecular data was annotated for type of molecular marker (mitochondrial versus nuclear data) and for the number of loci used.

2.2.3 Genetic discontinuities

To evaluate geographical discontinuities of genetic data across savanna taxa, we collected genetic patterns provided on research articles for genetic lineages with statistic support in both basal and crown nodes. For species with multiple inferences using different tree reconstruction methods, we gave priority to Bayesian inference, followed by Maximum likelihood and Maximum parsimony (Posada and Buckley 2004). The thresholds to account for node support were as follows: ≥ 95% for Bayesian posterior probabilities and ≥ 70% for bootstrap support (Alfaro, Zoller and Lutzoni 2003). For haplotypic networks and microsatellite analysis, lineages were accounted following author’s recommendations. Information provided by multiple articles on the same taxon was combined, when leading to congruent results. When in discordancy, priority was given to nuclear markers (microsatellite loci over sequences) (Fisher-Reid and Wiens 2011), and within the same type of marker, only the most recent research article on that taxon was considered. Whenever available, we recorded divergence time estimates for lineages according to the 95% confidence intervals reported. If multiple calibrations were available for the same taxa, we combined them to achieve a consensus of the estimated time intervals.

A spatial representation of statistically supported splits was made using QGIs v. 3.12 (QGIS Development Team 2021), by drawing lines in between genetic lineages. Lines very closely placed were overlapped to define patterns. This was performed in a conservative way, following authors' suggestions of potential barriers. A certain degree of sympatry between lineages was allowed, considering an overlap of up to 3 samples or sampling-points and drawing the line in a way that would exclude the minimum number of samples. In case of lack of statically supported clades for the analysed taxa or no geographical correspondence to statistically supported clades, no line was drawn, and this information was recorded as 'no patterns detected'. Each line in QGIS was associated with relevant attributes including research article, taxon name, taxa group and divergence time. To summarize the genetic patterns (Figure S2.1), we developed a new QGIS plugin named "linepattern" (see Text S2.2; Figures S2.2 and Figure S2.3 for details). Each simplified line is associated with the information of the number of lines used to generate it, the average, standard deviation, and range of the angles found, plus the ratio of number of lines per angle amplitude. We used a square size for line extraction of 0.5.

The resulting simplified patterns, hereafter referred as "relevant", were analysed per taxonomic group, and graduated according to the 'pretty breaks' mode in QGIS. We used the default number of classes and excluded the first class, assuming it represented the residuals for the cumulative patterns describing each group. Mitigating possible noise created from unequal sampling across groups, a different analysis was performed to summarize relevant patterns across all taxonomic groups. We created a new database, combining previous results on relevant patterns for each of the groups and re-ran the plugin. The same graduation method was used upon the results.

2.3 Results and Discussion

2.3.1 Number, timing, and diversity of available studies

From all research articles retrieved from the Web of Science, a total of 173 passed our filtering criteria (see Material and Methods section), and were, therefore, used to define patterns of genetic discontinuities (Table S2.2). From these, we were not able to detect any pattern on 30 research articles (see section 2.3.2 for more details). The first study considered in our dataset was from 1996 and dealt with the Smith's red rock rabbit (*Pronolagus rupestris*), based on the analysis of restriction fragment length polymorphisms (RFLP) of mtDNA (Matthee and Robinson 1996). Subsequent decade shows very few published articles per year; after 2006, we witness an increase of

published literature on this matter, for African sub-Saharan taxa, especially when considering in cumulative terms (Figure 2.1). However, disparities are evident in the selected molecular tools, taxonomic focus, and sampling.

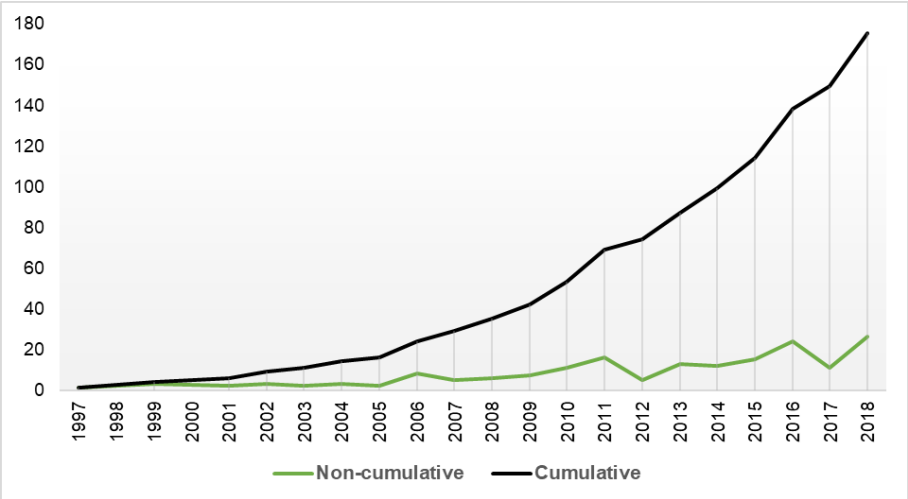


Figure 2. 1 Graph showing the number of research articles used in the analysis, per year of publication, both in non-cumulative (in green), as well as in cumulative terms (in black).

The majority of research articles in our dataset (38%) relied only in mtDNA sequences. This is not surprising as mtDNA was predominantly used on phylogeographic studies (Avise et al. 1987; Avise 1998, 2009) due to its simplicity to generate data and evolutionary peculiarities, including rapid evolutionary rates and lack of recombination (Rokas and Carroll 2005; Xia et al. 2009; Allio et al. 2017). However, inferences based solely on mtDNA can be misleading because of its maternal inheritance (Ballard and Whitlock 2004), and the integration of different molecular markers was also frequently retrieved in the analysed studies, mainly combining mtDNA sequences with nuclear sequence data (nuDNA) or microsatellite loci (31% and 4% of studies, respectively). Moreover, with the more recent advances in genomic tools, genome-wide studies are also available concerning our subject, but are still very scarce, accounting with only 3% of the research articles used in our database (Figure S2.4).

In terms of taxonomy, studies have largely been focused on vertebrate species (78% of retrieved articles), being mammals the most sampled group (47%), followed by reptiles (11%), birds (9%), amphibians (6%), and fish (5%). Invertebrates only accounted for 5% of the articles retrieved, whereas plants were the second most sampled taxonomic group, accounting for 17% of total articles in our dataset. These numbers reflect trends in the evaluation of phylogeographical patterns of taxa, but we acknowledge that our exclusion criteria may introduce some bias in this evaluation, according to the known range-body size relationship, as postulated by the macroecological theory (Brown and Maurer 1987;

Boyer and Jetz 2012). Nevertheless, the observed bias towards vertebrates seems a general issue within phylogeographic evaluations (Beheregaray 2008). Altogether, this review accounted with a taxonomic coverage of 286 species and 146 genera (Figure 2.2). The majority of studies have focused a single species (68%), but several have analysed species complexes and/or phylogenetic relationships among genera. A few comparative studies within specific geographic areas, such as in the West (Nesi et al. 2013; Hassanin et al. 2016; Huntley and Voelker 2016; Huntley et al. 2018) and in the East (Mairal et al. 2017; Barratt et al. 2018), as well as across the whole sub-Saharan Africa (Arctander, Johansen and Coutellec-Vreto 1999; Lorenzen et al. 2012; Bertola et al. 2016), are also focused on a single taxonomic group.

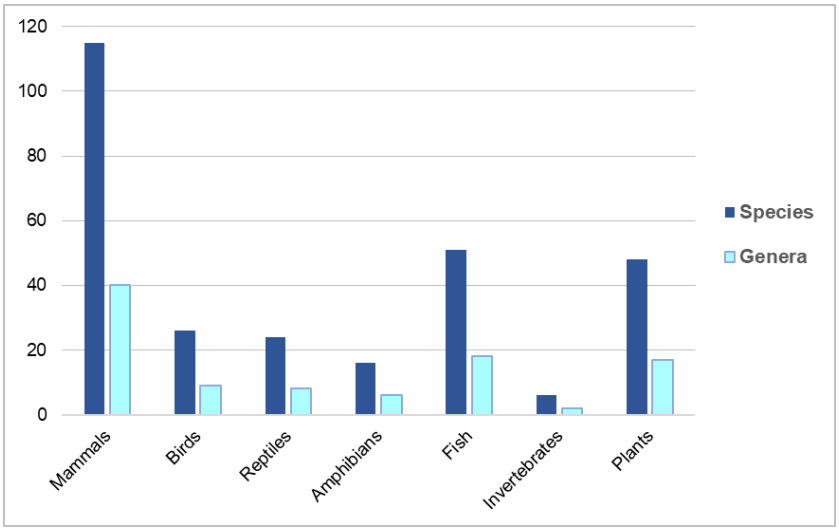


Figure 2. 2 Graph showing the number of species and genera registered, per taxonomic group.

Spatial sampling across sub-Saharan Africa varied among regions (Figure S2.5). Sampling across the three main regions (West, East, and South, as defined by Hewitt 2004) accounted for 14% of the total number of articles, whereas sampling covering only one or two regions, holds for 32% and 39% of total articles, respectively. Sampling covering the three regions, plus the Congolian Forest (as defined by Linder et al. 2012) accounted for 16% of total articles. Sampling was biased towards the Southern region, holding for the highest percentage across all comparisons (19%) (Figure S2.5). This is in line with highest sampled countries in the area – namely South-Africa – however did not reflected a higher number of relevant patters detected (see next section).

2.3.2 General description across the study-area

No patterns detected

As previously stated, for a total of 30 research articles on mammals, birds, reptiles, fish and plant species, no geographical patterns of genetic discontinuities were detected, either by lack of statically supported clades for the analysed taxa (n = 16) or by no geographical correspondence to statistically supported clades (n = 14). Examples of the first case report mainly species with high dispersion ability that enables high rates of gene flow, including, but not limited to, fruit bats that implement long-range annual migrations across the continent (Hassanin et al. 2016), habitat generalists with considerable dispersion ability (Russo, Chimimba and Bloomer 2006) or trees with long-distance wind dispersal of seeds (Sexton et al. 2015). Lack of statically supported clades is observed in studies exhibiting cyto-nuclear discrepancies (e.g. the African pygmy mouse, *Mus minutoides*; Bryja et al. 2014), life history discrepancies accounting mor for genetic differentiation than geographic vicariance (e.g. the rodents *Mastomys coucha*; Sands et al. 2015) or species-complexes with low levels of genetic differentiation and/or lack of monophyletic clades (e.g. the olive sunbird, *Nectarinia olivacea/obscura*; Bowie et al. 2004). Discrepancies between statistically supported clades and geographical distribution of species/populations are reported when diversity does not conform to existing taxonomic classifications (e.g. African green monkeys, *Clorocebus* spp.; Haus et al. 2013) or when species present overlapping ranges (e.g. the East African gerbils, *Gerbilliscus vicinus* and *G. nigricaudus*; Aghová et al. 2017).

Patterns of genetic discontinuities

A total of 19 genetic discontinuities were retrieved in our analysis across sub-Saharan Africa (Figure 2.3 and Figure 2.4). Our analysis clearly demarked the Tropical Forest belt (barrier I) and the Western Rift (II), which were previously used to define the main three regions by Hewitt (2004). According to this, and to facilitate interpretation, we numbered genetic discontinuities in our study in West (W), East (E) and Southern Africa (S). These were named after described geomorphological features or following our interpretation of most probable cause of discontinuity (based on climatic and biogeographic notions of the region, as well as author's descriptions). We retrieved six discontinuities within the West, six within the East, and five within Southern Africa. Most patterns revealed as relevant genetic discontinuities across several taxonomic groups (Figure 2.3), likely reflecting common evolutionary trajectories due to similar responses to biotic and abiotic factors.

Two genetic discontinuities in the Northwest region of Africa, the Dahomey-Volta (W2) and the Dahomey-Niger complexes (W3), were the most consistently retrieved across taxonomic groups considering the whole area, namely in mammals, birds, reptiles, and

plants. This is followed by the Cameroon Volcanic Line (W4) in the Western region as well, which showed relevant genetic discontinuities for the remaining three taxonomic groups (amphibians, fish, and plants) and parts of the Western Rift (II) for which different combinations of three taxonomic groups including mammals, reptiles, fish, and plants showed relevant genetic discontinuities (Figure 2.3 and Figure 2.4).

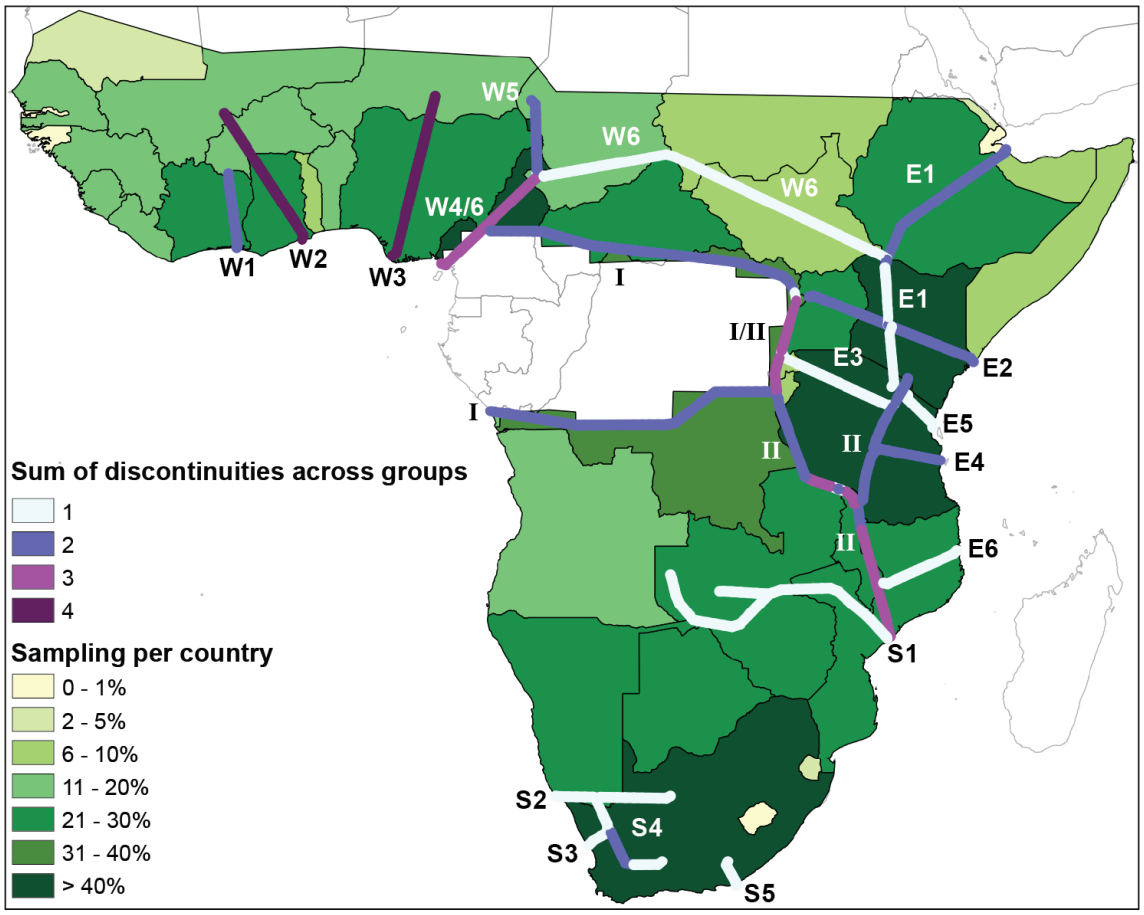


Figure 2. 3 Map summarizing relevant patterns across all taxonomic groups. Each of the 19 discontinuities is identified by a letter, according to region: W – West; E – East; and S – Southern Africa, and number code (see text and Figure 2.4 for details). Discontinuities are graduated according to the number of groups registering it as relevant. Also depicted is the study-area, according to a heat-map of sampled countries across all groups (percentage relative to total studies).

The Western Rift was not identified has a relevant phylogeographic discontinuity for birds and amphibian species, according to the articles review and applied criteria. Remaining retrieved genetic discontinuities, were considered relevant for only combinations of two different taxonomic groups, or exclusively detected in one taxonomic group (Figure 2.3 and Figure 2.4). Among the latter case, we observed three patterns only retrieved for mammals, named the Eastern transition zone (E3), the Zambezi-Kafue complex (S1) and the Bedford Gap (S5); two patterns only relevant for amphibian species, the Pagani (E5) and Lúrio rivers (E6); and two patterns exclusively observed for reptiles, the Orange River (S2) and the Knersvlakte (S3). The Central Africa Shear Zone

(CASZ) (W6) only emerged as a genetic discontinuity among fish in our literature search, which relates to its importance in geomorphological evolution of hydrological networks. It should be noted that other taxa were also reported for the genetic discontinuities presented, however did not gather a cumulative number of studies to be shown as "relevant" (see Material and Methods section). Nevertheless, examples of such cases will be given.

We were also interested to understand if the absence of a genetic discontinuity necessarily reflects its lack of importance for a taxonomic group, or rather a lack of studies across such taxonomic group in a region. For that, we estimated the ratio of studies in each country for each taxonomic group relative to the total number of studies for that taxonomic group, and plotted results as heat-maps. There is a sampling bias towards several countries across savanna biomes, with Kenya, Tanzania, Cameroon, and South-Africa receiving the highest research attention regarding the overall phylogeographical studies available (Figure 2.3). However, this general pattern is certainly biased by the highest number of studies on mammals, and different countries emerge for this ratio if each taxonomic group is considered separately (Figure 2.4). Still, we did not find a direct positive relationship between number of studies in a country and the number of times genetic discontinuities played a relevant role within a taxonomic group or among the different groups (Figure 2.3 and Figure 2.4). Explicit examples of this are the Dahomey-Volta (W2) and the Dahomey-Niger complexes (W3), retrieved as the most relevant when considering all taxonomic groups, and supported by 28 and 21 studies, respectively, which are not located across the most sampled countries, but in areas that have been averaged sampled across sub-Saharan countries (11% to 30% considering overall sampling of each taxonomic groups). We note that some genetic discontinuities considered as relevant in previous phylogeographic evaluations were not retrieved as relevant in our analysis. In some cases, this may be linked to methodological choices, as the exclusion of taxa with restricted distributional ranges. Namely, we acknowledge that genetic discontinuity named Eastern Rift (E1), bisecting Ethiopia towards Kenya, and argued as an important geographic barrier across several taxa, e.g. gelada baboon, *Theropithecus gelada* (Belay and Mori 2006), the Ethiopian wolf, *Canis simensis* (Gottelli et al. 2004), and several anuran species (Freilich et al. 2016), may have been less weighted in our database for these taxa due to the mentioned criteria.

Additionally, we observed a relationship between the number of studies within a specific taxonomic group and the number of genetic discontinuities retrieved for that specific group. Almost half of the articles in our database were focused on mammal species, which revealed the highest number of genetic discontinuities among taxonomic groups (n = 13) (Figure 2.4a).

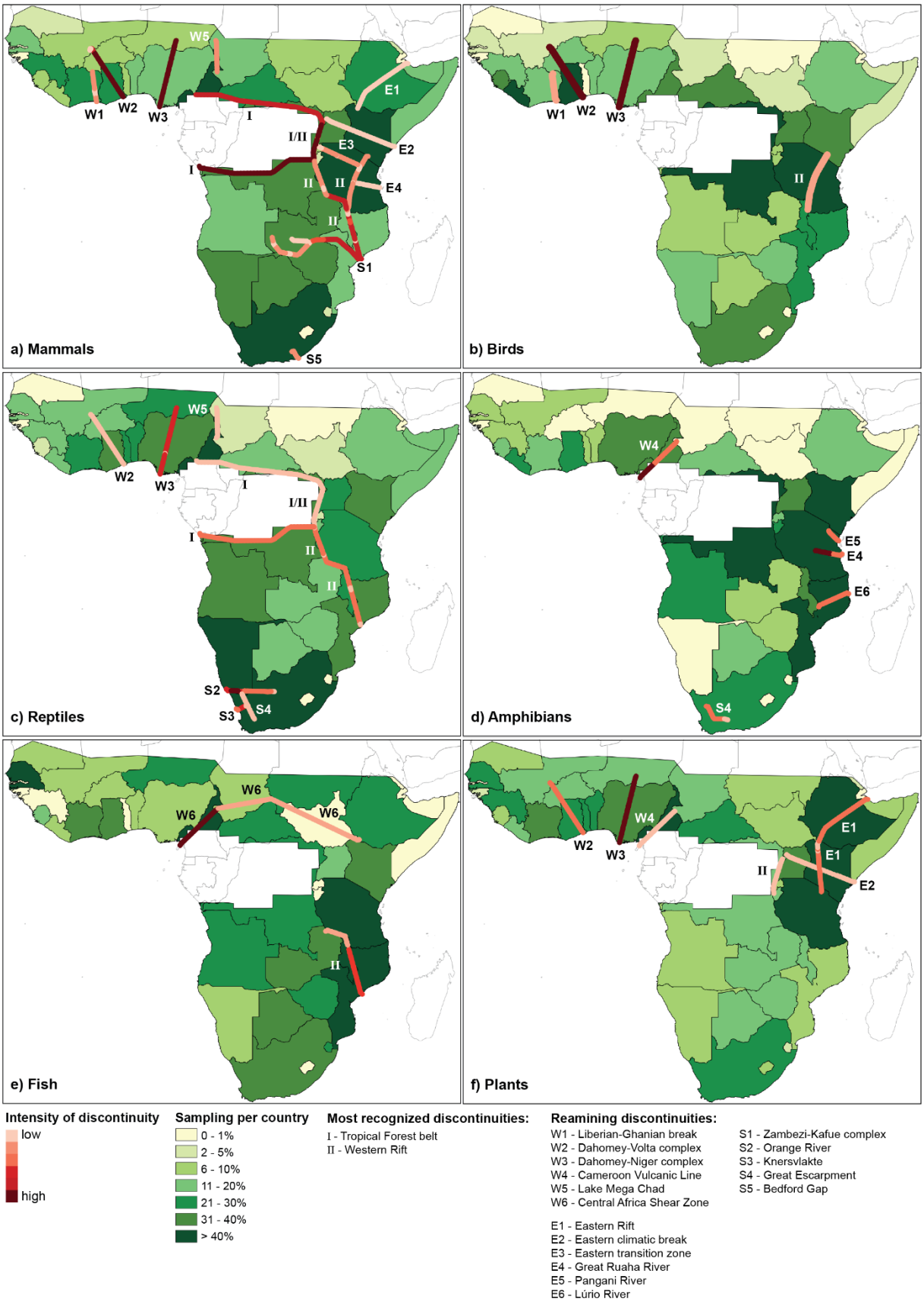


Figure 2. 4 Maps showing the resulting relevant genetic discontinuities across sub-Saharan Africa, per taxonomic group: **a)** Mammals; **b)** Birds; **c)** Reptiles; **d)** Amphibians; **e)** Fish; **f)** Plants. Each of the 19 discontinuities is identified by a letter and number code (following Figure 2.3 and the present legend) and its intensity is graduated according to the colour-

scheme in the legend. Also depicted is the study-area according to a heat-map of sampled countries for each group (percentage relative to total studies for each group).

Positioned on the other side of the curve, we could not retrieve from the literature consistent genetic discontinuities within invertebrates, certainly due to the very poor investigation of phylogeographic patterns for this taxonomic group. Notwithstanding, available information on invertebrates corroborates patterns found for other groups, including part of the Western Rift (II) (DeJong et al. 2003), and the Bedford Gap (S5) in South-Africa (Daniels et al. 2009) (Figure S2.1g).

The Pleistocene is one of the most important periods of diversification across sub-Saharan taxa, with 70% of retrieved divergence timing covering it (Figure 2.5), particularly for mammals (83%), amphibians (73%), and birds (100%, but corresponding to only two studies). Exceptions can be found for the CASZ (W6), Orange River (S2) and Knersvlakte (S3) discontinuities, as well as for reptiles in general, with 70% of the divergence events estimated before the Pleistocene. Fish are the group with the oldest divergence events, ranging from the Oligocene to the Pliocene periods (Figure 2.5).

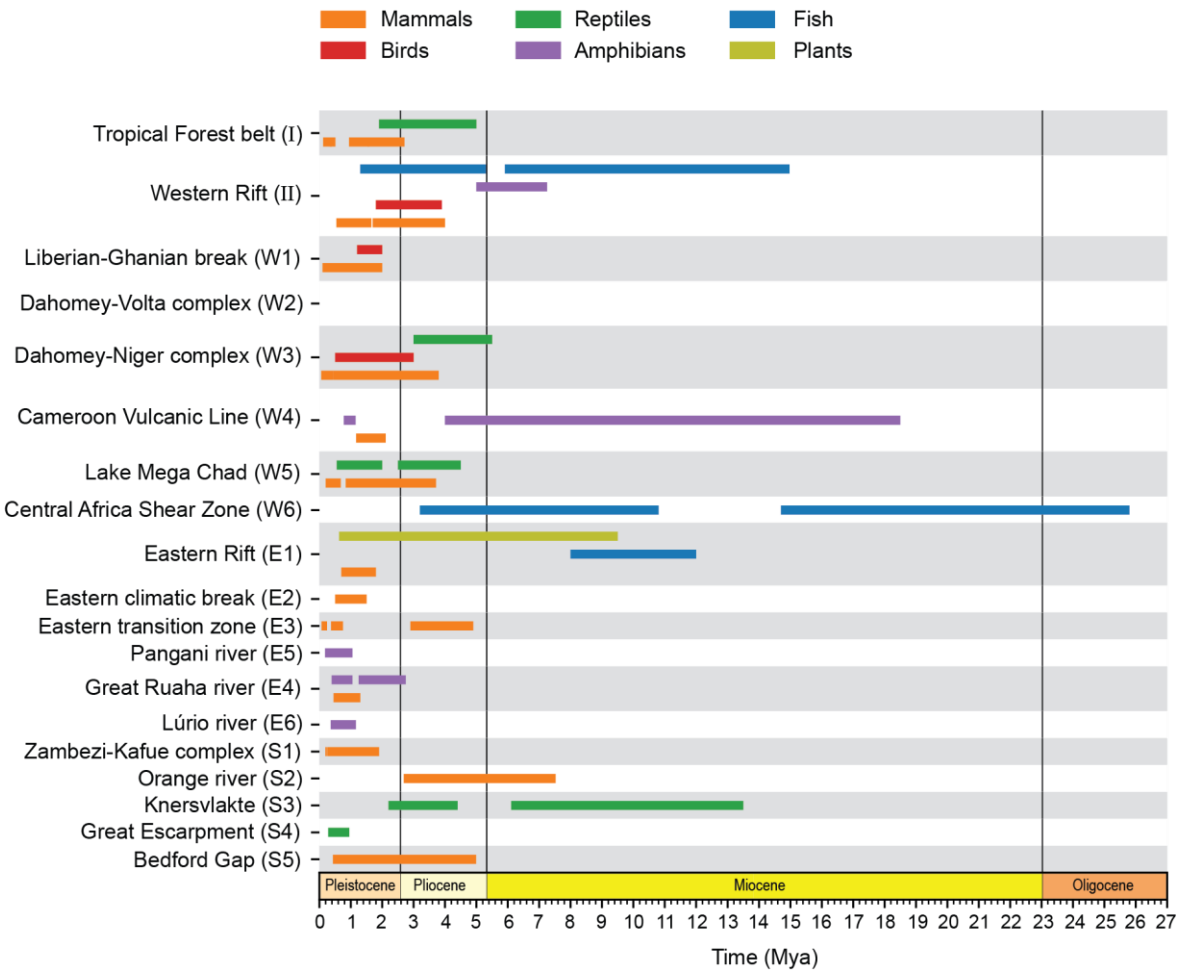


Figure 2. 5 Graph showing the retrieved divergence time estimates, according to genetic discontinuity and taxonomic group.

2.3.3 Genetic discontinuities imposed by the two most recognized features across sub-Saharan Africa

Tropical Forest belt

The Tropical Forest belt, across Central Africa, ranges from the west coast of Cameroon, Gabon, Congo Republic, and Democratic Republic of Congo, towards East Africa ending in the Albertine Rift, and has been associated with diversification events in different historical times since the Miocene until the Pleistocene climatic oscillations (DeMenocal 2004). In the last 2.5 Myr, 20 glacial cycles have promoted the alternating extension and contraction of forest habitats across this region (Hamilton and Taylor 1992; Dupont 2011). The presence of a continuous patch of tropical forest during pluvial periods has been seen as an important barrier for dispersal between the North and the South of this biological barrier, and has certainly driven divergence among savanna-associated taxa. We found this signature on the analysis of two taxonomic groups, the mammals retrieved from 14 articles, and the reptiles, using information from four articles. However, this pattern is also clear for birds (Marks 2010) and plants (Pock Tsy et al. 2009), but the lack of literature does not allow to retrieve it as consistent, as observed for mammals and reptiles.

Among several mammals, the Tropical Forest belt represents a distributional gap imposing divergence and resulting in different evolutionary clades present today in the north and in the south of the rainforest. We find this pattern in the literature for cheetah (*Acinonyx jubatus*), white-tailed mongoose (*Ichneumia albicauda*), common warthog (*Phacochoerus africanus*) (Muwanika et al. 2003; Dehghani et al. 2008; Charruau et al. 2011), among many others. Diversification in mammals have been apparently associated with a more recent period. For instance, the spotted hyena (*Crocuta crocuta*) exhibits an evolutionary split in North and South genetic clades associated with differentiated Pleistocene refugia and posterior expansions during the last 10 kyr (Rohland et al. 2005; Sheng et al. 2014), and both the lion (*Panthera leo*) and the hartebeest (*Alcelaphus buselaphus*) have divergence time estimates between northern and southern populations during the Pleistocene (0.1-0.5 Mya; Arctander et al. 1999; Flagstad et al. 2001; Bertola et al. 2016).

Across reptiles and plants, we found examples in the literature that set divergence events separating lineages in the north and in the south of the Tropical Forest belt to older periods in comparison with mammals, as is the example of the puff adder (*Bitis arietans*), which lineages diverged between 1.9 and 5.0 Mya, apparently related to

replacement of forests by more open habitats during the Miocene (Barlow et al. 2013). For savanna plant species, diversification events associated with the Tropical Forest are probably influenced by equatorial rainfall fluctuations during the Miocene towards the Pleistocene period, causing the expansion of populations towards south-east during periods of forest retraction, as well exemplified for the study of the African baobab (*Adansonia digitate*; Pock Tsy et al. 2009*)*.

Western Rift

The Western Rift discontinuity, as presented here, integrates both the western branch of the EARS, also known as the Albertine Rift (including the highlands across Rwanda, Burundi, Uganda, western Tanzania and DRC, all the way through Lake Malawi and into Mozambique), as well as the Eastern Arc Mountains (EAM), extending from south-eastern Kenya through the eastern half of Tanzania and southern Malawi (Chorowicz 2005). Remaining parts of the EARS, comprehending the eastern branch, will be discussed further, as they mostly represent differentiation within the eastern region (pattern E1). The Western Rift is a well know geomorphological barrier that separates Eastern from Southern Africa. Its influence in the genetic discontinuity of species and populations is generally explained by the interplay of tectonic activity driving geomorphological changes and climatic oscillations, with habitat fluctuations and placement of refugia. The EARS suffered increased tectonic activity during the Pleistocene, approximately 2 Mya (Chorowicz 2005; Roberts et al. 2012; Jess et al. 2020), when different fault basins were formed, making the Rift an important barrier to dispersal. However, the uplift of the southern section of the Rift began 8 Mya, during late-Miocene (Foster et al. 1997), while the uplift of the modern mountains of the EAM started around 7 Mya (Griffiths 2011). The altitudinal range of the mountains enabled long-term climatic stability against extreme drought (Lovett 1993), acting as a highland refugia, retaining diversity, and promoting diversification (Burgess et al. 1998, 2007).

On one side of the Western Rift extends the region known as Southern Africa, which went through stable climatic cycles during the Pleistocene, maintaining woodland and savanna habitats, probably fed by the Makgadikgadi paleolake across the Kalahari (Lancaster 1989; Shaw and Thomas 1996; Burrough, Thomas and Bailey 2009; Moore, Cotterill and Eckardt 2012). The region has been considered a long-lasting refugia for fauna and flora (Dupont 2011), due to the lower number of disclosed phylogeographic patterns compared to other African regions (Lorenzen et al. 2012). During the same period, Eastern Africa, on the other side of the Western Rift, experienced extreme dry periods leading to water-level reduction in the Great Lakes, which were converted into

closed saline systems (Verheyen 1996; Sepulchre et al. 2006; Stager and Johnson 2008; Lyons et al. 2015).

The western branch of the EARS was retrieved in our analysis as a genetic discontinuity for mammals, amphibians, and fish, either along its complete extension, or partially. Among mammals, rodents were the most well-represented group in the literature showing lineage sorting and allopatric phylogroups in the two sides of the Western Rift (Matthee and Robinson 1997; Nicolas et al. 2008; Castiglia et al. 2012; Colangelo et al. 2013; Bryja et al. 2018; Mazoch et al. 2018; Petružela et al. 2018). For other mammals, such as the sable antelope (*Hippotragus niger*), the Western Rift represents a sharp geographic barrier, which promoted vicariance of lineages likely due to the fragmentation of its habitat, the miombo woodlands (Pitra et al. 2002; Jansen van Vuuren et al. 2010). Springhare species of the genus *Pedetes* represent another example of vicariant lineages that likely suffered divergence driven by the influence of the Western Rift (Matthee and Robinson 1997). For several mammal and bird species, the Western Rift genetic discontinuity revealed to be contemporary, and eventually also driven by Pleistocene climatic oscillations (Arctander et al. 1999; DeJong et al. 2003; Zinner et al. 2009; Lorenzen et al. 2010; Miller et al. 2011; Castiglia et al. 2012). In the north of the Western Rift, an example from plants, namely by the African cherry tree (*Prunus africana*), represents the divergence between west and eastern populations, which has also been linked to divergent climatic conditions (Kadu et al. 2013).

Divergence time estimates for amphibians and fish (~0.5-1.5 Mya; Bittencourt-Silva et al. 2017; Sungani et al. 2017) are concordant with a tectonic event that led to the formation of the southern part of the Rift, during late-Miocene, separating Lake Malawi from the Rovuma River. This event changed the boundaries of the major drainage basins in the area and reversed the direction of several rivers (Goudie 2005; Stankiewicz and de Wit 2006), preventing gene flow for water-dependent taxa. During the same period, the uplift of the Livingstone Mountain also promoted vicariant events between Malawi and the Rufiji-Ruaha basins (Sungani et al. 2017). Divergence time estimates for fish genera *Hydrocynus* (Goodier et al. 2011), during Plio- and Pleistocene periods (1.3-5.3 Mya), can be linked to the uplift of the northern part of Lake Malawi.

Nowadays, the EAM comprehend isolated patches of montane forest habitats within savanna extensions, not representing an obvious barrier for animal dispersal (Petružela et al. 2018). However, it appears in our literature analysis as an important past geographic barrier promoting allopatric isolation in mammals and birds. Observed divergence is, however, mostly complemented by habitat contractions and expansions during Pleistocene climatic cycles that have induced genetic differentiation and lineage sorting. Examples among small mammals, including the red rock rat *Aethomys*

chrysophilus (Mazoch et al. 2018) and the Natal multimammate mouse *Mastomys natalensis* (Colangelo et al. 2013), evidence that the presence of unsuitable habitat, such as deep palaeolakes widespread in the region during humid periods (Trauth et al. 2005), may have contributed to the fragmentation and isolation of populations. Estimates of population divergence for Natal multimammate mouse (0.7-1.3 Mya), as well as for the split of the spiny mouse *Acomys ngurui* from the rest of its species complex (1.7-3.0 Mya), are concordant with such wetter phases of the Pleistocene (Colangelo et al. 2013; Petružela et al. 2018). Among birds, the *Myrmecocichla* genera offers a good example of vicariance driven by climatic oscillations in the EAM area, when forest expansion in wetter periods of the Pleistocene left less suitable habitats for arid-adapted species, promoting speciation in the genus (Voelker et al. 2012). Across afro-alpine species, two Pleistocene dispersal-migration scenarios are postulated: i) the “montane-forest bridge hypothesis”, stating that, during pluvial periods, the montane forest expanded to lower altitudes, facilitating short-range or stepping-stone dispersal between mountain ranges; ii) the “long-distance dispersal hypothesis”, stating that, during Pleistocene pluvials, phylogeographic patterns were mediated by long-distance dispersal between isolated mountain populations. Both hypotheses are non-mutually exclusive (Mairal, Sanmartín and Pellissier 2017; Chala et al. 2017).

2.3.4 Genetic discontinuities in savanna-associated taxa

Liberian-Ghanaian break (W1): It was retrieved for mammals (n = 5 studies) and birds (n = 2). No current geographic feature explains this genetic discontinuity; instead, the Pleistocene climate oscillations have been argued as the main driver of diversification in this region (Moodley and Bruford 2007; Nicolas et al. 2008; Bohoussou et al. 2015; Jacquet et al. 2015; Gaubert et al. 2016), in accordance with the estimates of divergence retrieved exclusively within this geological period (Figure 2.5). Detailed climatic and paleo-vegetation reconstructions have detected two putative forest refuges in Liberia and Ghana, which remained stable during several forest fragmentation stages (Maley 1996; Anhuf 2000). The diversification events in both forest and savanna-dwelling mammals, such as the tree pangolin (*Manis tricuspis*) (Gaubert et al. 2016), the Edward’s swamp rat (*Malacomys edwardsi*) (Bohoussou et al. 2015), the typical striped grass mouse (*Lemniscomys striatus*) (Nicolas et al. 2008) and the bushbuck (*Tragelaphus scriptus*) (Moodley and Bruford 2007), as well as forest-dwelling birds of genus *Criniger* (Huntley et al. 2018) and the green hylia (*Hylia prasine*) (Marks 2010), are concordant with the vicariance scenario between both forest refuges. In addition, the Mascarene grass frog (*Ptychadena mascareniensis*), inhabiting both humid savannas and open forests, also

shows lineage sorting coincident with this break, although representing divergence events during the Pliocene (Zimkus et al. 2017). This suggests that older climatic changes may still influence current distribution patterns across taxa in this region (Zimkus et al. 2017). Interestingly, the Liberian-Ghanaian break was also retrieved as genetic discontinuity for the savanna-dwelling Guinea multimammate mouse (*Mastomys erythroleucus*), with vicariance associated with the fragmentation of savanna habitats during humid periods, but also to the Sourou and Volta rivers (Brouat et al. 2009). In this case, the Liberian-Ghanaian break probably represents a contact zone for the sampled phylogroups, rather than a break for forest-dwelling species.

Dahomey-Volta complex (W2) and Dahomey-Niger complex (W3): These two genetic discontinuities represent the Dahomey Gap, a large, dry, open area, currently bounded by the Volta River to the west, and the Wemé River to the east. The Dahomey Gap has ca. 3 Myr (Robert and Chamley 1987) but the extension and vegetational composition of the area has varied throughout the Pleistocene and Holocene periods, being occupied by forest habitats at least between 115 kya and 129 kya, and between 4.5 kya and 8.4 kya (White 1983; Dupont et al. 2000; Salzmann and Hoelzmann 2005). This enabled several forest-dwelling taxa, with some degree of ecological tolerance for savanna-forest mosaics, to persist, adapt, or even expand their distribution during drier conditions, increasing the level of endemism of this area (Colyn et al. 2010; Nicolas et al. 2011; Gaubert et al. 2016; Voelker et al. 2017). Nevertheless, this vegetational dynamics has been considered in vicariance patterns together with the position of both Volta and Niger rivers (Booth 1958; Robbins 1978; Couvreur et al. 2020). Dahomey-Volta complex and Dahomey-Niger complex were retrieved in our literature analysis for mammals (n = 11 and n = 13 studies, respectively), birds (n = 4 studies for both), reptiles (n = 2 and n = 4, respectively), and plant species (n = 3 and n = 5, respectively). Taxa related to these genetic discontinuities can either i) be endemic to the Volta-Niger interflow, as seen in mammals (Nicolas et al. 2008; Colyn et al. 2010; Nicolas et al. 2011; Gaubert et al. 2016), birds (Voelker et al. 2017), and plants (Demenou et al. 2016, 2018); ii) present disjunct distribution between Dahomey-Volta and Dahomey-Niger complexes, as recorded for mammals (Nesi et al. 2013; Hassanin et al. 2015, 2016), birds (Marks 2010; Huntley and Voelker 2016; Huntley et al. 2018), reptiles (Eaton et al. 2009; Leaché and Fujita 2010), amphibians (Jongsma et al. 2018), and plants (Koffi et al. 2011; Daïnou et al. 2014); or even iii) be separated either by the Dahomey-Volta complex, as retrieved for mammals (Nicolas et al. 2011; Bryja et al. 2014; Jacquet et al. 2015), and birds (Huntley and Voelker 2016; Fuchs, Fjeldså and Bowie 2017), or by the Dahomey-Niger complex, as seen for mammals (Alpers et al. 2004; Nicolas et al. 2011; Colangelo et al.

2013; Bertola et al. 2016; Lobon et al. 2016), reptiles (Trape et al. 2009), and plants (Duminil et al. 2013).

The savanna corridor within the Dahomey Gap has been discussed as an effective barrier for several strict forest-dwelling species, preventing dispersal and leading to currently disjunct distribution (Eaton et al. 2009; Leaché and Fujita 2010; Marks 2010; Koffi et al. 2011; Nesi et al. 2013; Daïnou et al. 2014; Hassanin et al. 2015, 2016; Huntley and Voelker 2016; Huntley et al. 2018; Jongsma et al. 2018). Allopatric differentiation has been associated to aridity peaks during the Pliocene (Hassanin et al. 2015; Huntley and Voelker 2016; Huntley et al. 2018) and the Pleistocene periods (Nesi et al. 2013; Huntley et al. 2018), supporting the role of forest contractions as the main driver of diversification in this region. Some diversification events of endemic lineages from the Volta-Niger interflow pre-date the Last Glacial Maximum (LGM) period (Nicolas et al. 2008; Gaubert et al. 2016; Demenou et al. 2018), others have occurred during the post-LGM (Demenou et al. 2016), reinforcing the role of successive forest expansion and contractions in generating local diversity (Hardy et al. 2013; Couvreur et al. 2020). On the other hand, generalist and/or savanna-dwelling species show little differentiation across the Dahomey gap, suggesting that the evolutionary role of this barrier is impaired by taxa ecological traits (Moodley and Bruford 2007; Jongsma et al. 2018). The Volta and Niger rivers may have acted as physical barriers, separating lineages of forest-dwelling (Nicolas et al. 2011; Bohoussou et al. 2015; Lobon et al. 2016), savanna-dwelling (Alpers et al. 2004; Nicolas et al. 2008; Brouat et al. 2009; Bertola et al. 2016), as well as generalist species (Colangelo et al. 2013; Fuchset al. 2017). Both rivers may also represent the distribution limit of re-expansions from refugial areas (Nicolas et al. 2011; Bohoussou et al. 2015).

Cameroon Volcanic Line (CVL) (W4): It is a 1,600 km long succession of volcanoes and plateaus aligned in an east-northeast/west-southwest direction, extending from the Gulf of Guinea to the Adamawa Plateau, in continental Africa (Ubangoh, Pacca and Nyobe 1998; Marzoli et al. 2000). The CVL has been suggested to have acted as both refugium (Maley 1996; Hardy et al. 2013; Jongsma et al. 2018; Migliore et al. 2019) and geographic barrier for several taxa, being the division between Guinea and Congo sub-biogeographical regions, as defined by Linder et al. (2012). It has been retrieved in our literature analysis as a relevant genetic discontinuity for amphibians (n = 3 studies; Evans et al. 2004; Portik et al. 2017), and plants (n = 2; Allal et al. 2011; Lompo et al. 2018), but also found as a divergence driver in a mammal study (Nesi et al. 2013). However, the volcanic activity does not seem to be the primary driver of diversification, since the main peaks of the CVL were established long before (between 7 and 32 Mya; Marzoli et

al. 2000) the timing of divergence estimated for several taxa, namely about 1.17–2.11 Mya for bat species of the genus *Megaloglossus* (Nesi et al. 2013); 0.78–1.15 Mya for the frog species *Scotobleps gabonicus* (Portik et al. 2017); and 0.92 kya for the African locust bean tree (*Parkia biglobosa*) (Lompo et al. 2018). Instead, lowland forest species with low vagility would have diverged in isolation within the refugia postulated in this area (Maley 1996), and the high-elevation volcanic peaks further restricted gene flow between adjacent populations (Portik et al. 2017). In addition, evidence for the refuge in CVL is indicated by the high genetic diversity, cryptic, and relict lineages, as well as high endemic species richness found in this area (Missoup et al. 2009; Hardy et al. 2013; Jongsma et al. 2018; Migliore et al. 2019).

Lake Mega Chad (W5): Occupying a central position across Northern Africa, the contemporary Lake Chad is considered a relic of a much larger paleolake, the Mega Chad. Although strongly debated, geomorphologic evidence confirms the recurrent cycles of transgression and regression of this paleolake (Durand 1982; Ghienne et al. 2002; Schuster et al. 2005; Griffin 2006). Since its origin during late-Miocene, between 5.0 Mya and 6.5 Mya (Griffin 2006), at least three lacustrine episodes were confirmed to have occurred during the late-Pliocene (at 3.0–3.5 Mya) and Holocene (around 8.0–10.0 kya and 4.4/5.4–7.0 kya) (Schuster et al. 2005). During humid periods, the lake would have reached more than 350 thousand km^2 (Ghienne et al. 2002; Schuster et al. 2005; Griffin 2006), probably creating a large geographic gap and promoting allopatric differentiation for savanna-dwelling taxa in this region. In our literature analysis, this discontinuity was retrieved as relevant in mammals (n = 7 studies) and in reptile species (n = 2 studies). These taxonomic groups were represented by the African grass rat, *Arvicanthis niloticus* (Dobigny et al. 2013), the African multimammate rat, *Mastomys erytroleucus* (Brouat et al. 2009), and vipers of the genus *Echis* (Pook et al. 2009). This discontinuity was also seen for the African baobab tree (*Adansonia digitata*) (Pock Tsy et al. 2009). Divergence time estimates associated with this genetic discontinuity encompass a large period, ranging from the Pliocene to middle-Pleistocene (Figure 2.5), marked by multiple cycles of expansion and contraction of savanna biomes (Hewitt 2004). Although the latest known transgression of Lake Chad, occurring over a relatively short period during the Holocene (Schuster et al. 2005), could not be related to the genetic differentiation of some taxa (Allal et al. 2011), the oldest known expansion, during late-Pliocene, is coincident with the divergence time estimated for the African grass rat (0.84–3.71 Mya; Dobigny et al. 2013), as well as for the vipers of the genus *Echis* (2.5–4.5 Mya; Pook et al. 2009). During dry periods, the Lake would recede, with the expansion of savanna and woodlands promoting population dispersal (Pook et al. 2009)

and secondary contact between previous isolated lineages (Moodley and Bruford 2007; Dobigny et al. 2013). However, caution must be taken when doing such associations, since large confidence intervals in time estimates preclude robust conclusions (Dobigny et al. 2013). Similar pattern of differentiation along this area displayed by several taxa have been solely attributed to climatic factors (Flagstad et al. 2001; Moodley and Bruford 2007; Zinner et al. 2009; Barlow et al. 2019).

Central Africa Shear Zone (W6): The uplift of the Central African Rift system, also known as the Central Africa Shear Zone (CASZ), started during the Cretaceous. Crossing Central Africa, from the Indian to the Atlantic Ocean, it only remains active nowadays across the Cameroon Volcanic Line (Guiraud and Maurin 1992; Stankiewicz and de Wit 2006). However, it was retrieved as an important genetic discontinuity in our analysis for fish, namely for genera *Synodontys* and *Hydrocynus* (Goodier et al. 2011; Pinton et al. 2013). It corresponds to the drainage separation of the Congo basin in Central Africa from other basins in Northern Africa, such as the Nilo-Sudan, Niger and Sanaga, mediated by tectonic activity (Goodier et al. 2011; Pinton et al. 2013). The CASZ presents the oldest diversification event recovered in our study (Figure 2.5), most likely acting as a physical barrier to dispersal between fish clades during late-Oligocene and Miocene (14.7-25.8 Mya) for *Synodontys*, and in the late-Miocene and Pliocene (3.2-10.8 Mya) for the *Hydrocynus* complex. Other studies on *Coptodon* and *Sarotherodon* fish genera partially corroborate this pattern, showing vicariance associated with the complex evolution of hydrological landscape in the remaining active part of the CASZ (Kide et al. 2016). Such studies support the idea that cichlid diversification is not only driven by ecological speciation (Burress 2015), with vicariance acting as a potentially important, but still neglected, evolutionary force (Kide et al. 2016).

Eastern Rift (E1): Representing the eastern branch of the EARS, also known as Gregory Rift, runs from the Afar triangle into southeast Kenya and results from the splitting process of the Nubian and Somali tectonic plates (Chorowicz 2005). Topography and land formation of the Rift underlie the genetic discontinuities detected for numerous taxa, with vicariant events dividing populations situated on either side. The Eastern Rift also represents the division between Ethiopian and Somalian biogeographical regions (Linder et al. 2012). This genetic discontinuity was retrieved both in mammal (n = 3 studies) and in plants (n = 3), but was also found for other taxa, namely birds and fish. For both the giraffe (*Giraffa camelopardalis*) and ostrich (*Struthio camelus*), this area broadly separates northern from eastern populations (Miller et al. 2011; Winter, Fennessy and

Janke 2018). For bushbuck (*Tragelaphus scriptus*), the divergence observed between the two main haplogroups was according to the Rift Valley axis, separating the species into Western-Northern and Eastern-Southern groups (Moodley and Bruford 2007). Across the large-eared free-tailed bat (*Otomops martiensseni*), this area is associated with the split of two reciprocally monophyletic clades, comprising the northeast, and remaining species distribution (Lamb et al. 2008). For the Nile tilapia (*Oreochromis niloticus*) it corresponds to the division between two ichthyofaunal clades (Bezault et al. 2011) and for several species of plants, it is a barrier dividing different afro-alpine populations (Wondimu et al. 2014; Chen et al. 2015; Gizaw et al. 2016; Mairal et al. 2017). Tectonic activity associated with the Rift uplift, combined with Plio-Pleistocene climatic cycles, would contribute to habitat instability in this area. Estimated divergence time between Nile tilapia's clades (8.0-12.0 Mya; Bezault et al. 2011) seems to be related to the tectonic disruption that led to the isolation of the Awash and Nile basins (Roberts 1975). Within the Awash system populations, low genetic diversity and high differentiation suggests strong effect of genetic drift within fragmented populations from isolated highland lakes (Bezault et al. 2011). The allopatric division between the two main bushbuck lineages is highly associated with geomorphology (Moodley and Bruford 2007) and deep divergence times recovered for both plant species, *Haplocarpha rueppelii* (Chen et al. 2015) and *Canarina eminii* (Mairal et al. 2017), between 630 kya and 9.5 Mya, suggest long-lasting vicariant events among populations on west and east of the Rift. Major faults and flooding areas associated with tectonic activity, may have isolated geographically close populations. Through adaptative evolution, such diversification mechanisms can lead to differences in behaviour and habitat preference between lineages. As in other regions of the Rift, Pleistocene interpluvials would benefit the presence of temperate-adapted taxa, whereas pluvials would restrict the distribution of such lineages to refugial zones. The close association found between bushbuck haplogroups and several core habitats attests for the persistence of these refugia throughout Pleistocene climatic cycles (Moodley and Bruford 2007). The pattern observed by Lamb et al. (2008) for bats, demonstrates the importance of past climatic fluctuations in shaping current taxa distribution, even across highly mobile species, with flight capability. Feeding mostly on insects and depending on cave colonies for roosting, the northeast clade, restricted by the Eastern Rift, probably suffered population contractions and even local extinctions during arid periods of the Pleistocene. The estimated divergence time between the northeast clade and remaining species distribution (0.7-1.8 Mya; Lamb et al. 2008) is coincident with a period of continental shift towards dryer and cooler temperatures (DeMenocal 2004), which could have forced bat populations to move towards lower altitude valleys, with limited habitat availability. For

plants like *C. eminii*, as well as *Carduus schimperi* (Wondimu et al. 2014), stepping-stone-dispersal through grasslands and open forest corridors might have occurred between mountains during pluvials, with subsequent *in situ* diversification during interpluvials. For *H. rueppelii*, as well as for *Carex monostachya* (Gizaw et al. 2016), vicariant events might have been ameliorated by long-distance dispersal via strong winds, bird ingestion, or mud on feet of mammals that would migrate between mountains during pluvials.

Eastern climatic break (E2): The Horn of Africa is the easternmost extension of Africa, including most of South-Sudan, Ethiopia, Somalia, and north-eastern Kenya. A combination of African, Mediterranean, Asiatic and ancient Gondwanan influences have shaped the flora and fauna of this region (Schreck and Semazzi 2004; Hildebrand et al. 2019). The resulting biota contains many arid-adapted species, mainly from the Ethiopian Ogaden Desert and northeast Kenyan semi-deserts (Soultan, Wikelski and Safi 2020). On the other hand, the Ethiopian highlands, the largest continuous area of mountains in the continent, have a tropical monsoon climate, with high mountains catching the precipitation from the Indian Ocean (Schreck and Semazzi 2004; Hildebrand et al. 2019). This diversity of climates in this region seems to create a genetic discontinuity for numerous taxa, that was retrieved in our analysis for mammals (n = 3 studies) and plants (n = 2), but it was also viewed for other taxonomic groups. For example, it separates cheetah subspecies *Acinonyx jubatus soemmeringii* and *A. j. raineyi* (Charruau et al. 2011), reticulated (*Giraffa camelopardalis reticulata*) from the Masai giraffe (*G. c. tippelskirchii*) (Winter et al. 2018), and differentiates clades for the sub-Saharan leopard tortoise, *Stigmochelys pardalis* (Fritz et al. 2010). Reported low rates of migration and gene flow between clades leads to a scenario of repeated isolation during climatic cycles, ultimately resulting in allopatric divergence of populations. This was the case for the African fossorial root-rat (*Tachyoryctes splendens*), with reciprocally monophyletic clades present on either side of this barrier, and divergence estimations dated to mid-Pleistocene (0.5-1.5 Mya; Šumbera et al. 2018). Having low-dispersion ability and preference for open moist habitats, extant root-rats' distribution is limited to humid highlands of Ethiopia in the north, and Kenya in the south (Jarvis 2013). Periods of dry climate, as well as extensive growth of forest cover during humid periods, would probably fragment root rat's suitable habitat, limiting it to refugia, from where they have recently expanded. The same has apparently followed for the Nutt's river frog (*Amietia nutti*), the heather plant (*Erica trimera*) and the giant lobelia (*Lobelia giberroa*), with significant allopatric division between Ethiopian and remaining southern populations (Kebede et al. 2007; Gizaw et al. 2013; Manthey et al. 2017).

Eastern transition zone (E3): This relevant genetic discontinuity separates northern from central Tanzania, and was only retrieved as relevant among mammals studies (n = 5). This region is topographic and climatically diverse, flanked by the Great Lakes and the EAM, being the transition zone between the Somalian and Zambezian biogeographical regions (Linder et al. 2012). Historical interplay between geographic features and climatic fluctuations might have contributed to genetic discontinuities, acting at both intra- and inter-specific levels. It is related, for instances, with the allopatric distribution of Mearns's pouched mouse (*Saccostomus mearnsi*) and the South African pouched mouse (*S. campestris*) (Mikula et al. 2016). Their diversification events took place during early Pliocene (2.9-4.9 Mya), a period from where faunal fossil records recognize the existence of both biogeographical regions (Mikula et al. 2016). For hartebeest (*Alcelaphus buselaphus*), this region is where the reciprocally monophyletic eastern and southern lineages partially overlap (Arctanderet al. 1999; Flagstad et al. 2001) and for the spotted hyena (*Crocuta crocuta*) it divides between both extant African lineages (Rohland et al. 2005; Sheng et al. 2014). The same barrier separates two distinct eastern clades for both the lion (*Panthera leo*) and the typical striped grass mouse (*Lemniscomys striatus*), linking geology and vegetation distribution (Nicolas et al. 2008; Bertola et al. 2016). This Eastern transition zone also relates to the clustering scheme presented by several Afroalpine species of plants, where populations from distant mountain massifs are genetically closer (e.g. *Hagenia abyssinica* reported by Gichira et al. 2017). Displayed genetic structure is probably linked to remnant refuge populations in the mountains, during the Pleistocene. Estimated divergence times leading to vicariant events for both the hartebeest (410-580 kya) (Arctanderet al. 1999; Flagstad et al. 2001), and the typical striped grass mouse (373-746 kya) (Nicolas et al. 2008), are coincident with periods of climate warming – pluvials. Increased temperature and humidity would lead, not only to the contraction of suitable savanna habitat, but also to the rise of the water level for several lakes. There is evidence for the occurrence of deep lakes in the Magadi-Natron-Olorgesailie basin around 1 Mya, as well as recorded alternation between lacustrine and subaerial environments for several drainages in the area, between 500 kya and 1 Mya – e.g. drainage reversal of Lake Kyoga and connection between Lake Victoria and Lake Edward drainages (Goudie 2005; Trauth et al. 2005). On the contrary, estimated time for the divergence between both lion clusters in this region (240-72 kya) (Bertola et al. 2016) is related to two dry periods – interpluvials. Dramatic habitat changes due to alternating cycles of hydrological extremes, could have acted as a periodical barrier between populations, which would persist within refugia of suitable habitat.

Great Ruaha (E4): Is a river in south-central Tanzania that flows through the Usangu wetlands and the Ruaha National Park, into the Rufiji catchment system. It has been proposed that large rivers and their flood plains can be important limiting factors for small-bodied mammals, such as rodents and elephant-shrews. The influence of rivers is probably related to their dwelling nature and associated shelters, leading to limited dispersal ability (Rathbun 2009; Coals and Rathbun 2013; McDonough et al. 2015; Mazoch et al. 2018). The Great Ruaha River seems to be a relevant phylogenetic discontinuity for both mammals (n = 3 studies) and amphibians (n = 2 studies). It is associated with the restricted distribution of the rufous elephant shrew (*Rhynchocyon petersi*) (Carlen et al. 2017), as well as the separation between Nguru spiny mouse's (*Acomys ngurui*) haplogroups in this area (Petružela et al. 2018). Rivers can also influence the genetic structure of amphibians. If, on the one hand, they can facilitate gene flow within drainages, due to downstream movement of eggs and/or larvae, a river can also act as barrier to dispersal due to general low individual mobility (Measey et al. 2007). Phylogenetic relationships within the East African reed frog (*Hyperolius substriatus*) were best-explained by minor hydrobasins that increase genetic structure at a finer-scale (Lawson 2013), specifically the Great Ruaha River, preventing gene-flow between Udzungwa and Usambara lineages (Bittencourt-Silva et al. 2017). The same barrier is coincident with the division of both central populations of two species of spiny reed frogs (*Afrixalus fornasini* and *A. delicatus*) (Barrat et al. 2018). Estimated divergence time for the split within Nguru spiny mouse (0.45-1.30 Mya) is coincident with described pluvial periods during the Pleistocene, where savanna habitats would be reduced to refugia, leading to allopatric populations, separated by unsuitable, forest habitat. Estimated divergence timing for the split within the East African reed frog (1.25-2.75 Mya) (Bittencourt-Silva et al. 2017) and the spiny reed frog (*A. fornasini*) (3.90-5.60 Mya) (Barrat et al. 2018) can, on the other hand, be related to a gradual aridification of East Africa during this period (Trauth et al. 2009). Given the fact that amphibians are water-dependent animals, drier conditions would lead to a loss of suitable habitat and promotion of vicariant events (Kissling et al. 2012).

Pangani River (E5) and Lúrio River (E6): These genetic discontinuities were only recorded as relevant for amphibian species, reported by Barrat et al. (2018). The Pangani River runs from northern Tanzanian highlands into the northeast coast of the country and is a barrier for northern populations of both the delicate spiny reed-frog (*Afrixalus delicatus*) and the dwarf squeaker (*Arthroleptis xenodactyloides*). The Lúrio River runs from mount Malema, in northeastern Mozambique, into the Indian Ocean and

is a barrier for the Mozambique populations of both the Fornasini spiny reed-frog (*Afrixalus fornasini*) and the yellow-spotted tree frog (*Leptopelis flavomaculatus*). The riverine barrier hypothesis states that rivers can act as dispersal barriers for numerous taxa, leading to vicariant events that precipitate allopatric speciation (Gaston 2000; Moritz et al. 2000; Voelker et al. 2013). Due to their poor dispersal abilities and highly structured demes, amphibian populations can experience such speciation events. For both generalist species of spiny reed-frogs (*Afrixalus delicatus* and *A. fornasini*), as well as for the forest-adapted yellow-spotted tree frog (*Leptopelis flavomaculatus*), demographic analysis by Barrat et al. (2018) were consistent with models of allopatric divergence by both rivers. For the forest-adapted dwarf squeaker (*Arthroleptis xenodactyloides*), demographic models were associated with historical gene-flow, most likely trough riparian corridors that facilitated dispersal between drainage basins across different mountains, during humid favourable periods (Zimkus et al. 2017). Divergence time estimates of the splits associated with the Pangani River (180-800 kya), as well as the ones associated with the Lúrio River (0.36-1.60 Mya) (Barrat et al. 2018), all lead to a period of climate warming during the Pleistocene era. During this period, described amphibian species would have persisted only in humid, forest refugia, and once suitable conditions arose, further expansion was limited by river barriers.

Zambezi-Kafue complex (S1): This feature corresponds to an approximated south/east division across Southern Africa, following both river courses, between Zambia, Zimbabwe, and Mozambique. It is also related with a stripe of tropical rainforest (Hewitt 2004). Studies describing this genetic discontinuity suggest that the river system does not constitute an insurmountable barrier *per se*, since most of the observed taxa have wider distributions across the continent. Rather, observed patterns result from an interplay between the cyclic climatic and habitat oscillations during the Pleistocene, and the geomorphological history of the region. In fact, the Zambezi River only established its current flow very recently (Goudie 2005). The Upper Zambezi, including the modern Kafue River, is thought to have evolved as a separate system from the Middle and Lower Zambezi, which became linked by tectonic activity during the Pliocene to middle-Pleistocene. This tectonic activity also led to course changes of several tributaries, as well as to the formation of deep escarpments in the middle Zambezi (Goudie 2005). The catchment of several rivers during the Pleistocene eventually resulted in the formation of extensive wetlands, including a series of floodplains and lakes in the Upper Zambezi and Kafue areas. It is therefore not surprising that many of the taxa associated with this genetic discontinuity are savanna-associated species. Both river systems probably acted as barriers to gene flow during climatic cycles for the last 2.5 Myr, leading to multiple

events of population contraction, isolation and expansion, resulting in complex biogeographical histories for these species.

This complex was considered a relevant genetic discontinuity only for mammals (n = 11 studies), but killifishes were also retrieved as supporting genetic discontinuities associated with Zambezi-Kafue river system (Bartáková et al. 2015). In the case of the South African pouched mouse (*Saccostomus campestris*), the Zambezi and Kafue rivers are described as current barriers to effective gene flow between the proposed subspecies *S. c. campestris* and *S. c. mashonae* (Mikula et al. 2016). The divergence time estimation (0.9-1.9 Mya) (Mikula et al. 2016) is related with climatic oscillations and intermittent geomorphological features, such as the Chambeshi–Kafue–Zambezi paleolake, that contributed to the fragmentation of ancestral populations. The suitable habitat for this species, composed of woodlands, grasslands, and scrubland savanna, was probably fragmented during the last pluvial period. Population expansion after savanna woodland habitats resurged, in the last 80-85 kyr, bringing both clades into contact, around current placement of the river system. Today, it continues to act as a permeable barrier to gene flow for this species (Mikula et al. 2016). The megadrought of the Chambeshi–Kafue–Zambezi paleolake, during drier phases of the Pleistocene, also influenced the isolation of populations of the bushveld gerbil (*Gerbilliscus leucogaster*), causing genetic drift (McDonough et al. 2015). Transitional zones along the Kafue flats were detected not only for this species, but also in the red rock rat (*Aethomys chrysophilus*) and the South African pouched mouse (*Saccostomus campestris*), reinforcing the importance of this area for the divergence of species (McDonough et al. 2015; Mikula et al. 2016; Mazoch et al. 2018). Major geomorphological reconfigurations of the Zambezi River have also been associated with vicariance events that originated a so-called “hotspot of karyotpical diversity”, comprising several chromosomal races of *Fukomys* mole rats (van Daele et al. 2007). The authors explain that water-level fluctuations in the Zambezi River, associated with climatic oscillations and a suitable dispersal habitat, triggered allopatric and peripatric speciation, both between and within phylogroups (van Daele et al. 2007). The same combination of geomorphological features acting as geographical barriers, and climatic oscillations promoting genetic discontinuities in this area, was also observed for the spiny mouse (*Acomys spinosissimus*) (Petružela et al. 2018).

The Zambezi-Kafue complex is also reported as a genetic discontinuity for several large-bodied mammals, widely distributed across savanna habitats, such as the hartebeest (*Alcelaphus bucelaphus*) (Arctander et al. 1999; Flagstad et al. 2001), the wildebeest (*Connochaetes taurinus*) (Arctander et al. 1999), the plains zebra (*Equus quagga*) (Pedersen et al. 2018), the giraffe (*Giraffa camelopardalis)* (Winter et al. 2018)

and the baboon (*Papio* spp.) (Zinner et al. 2009). The alternating pluvial and interpluvial cycles of the Pleistocene are pointed as the main cause of diversification for most species. For the hartebeest, divergence time estimations range between 190 and 240 kya (Flagstad et al. 2001), whereas for baboons, it ranges between 600 kya and 1.03 Mya (Zinner et al. 2009). The origin of some of these species seems to reflect the importance of the South as a refugia (Arctander et al.1999; Zinner et al. 2009), namely for the plains zebra, with its likely location on the Zambezi wetlands, an area of climatic stability throughout geological time (Pedersen et al. 2018).

Orange River (S2): This pattern results from a north/south genetic discontinuity along the western part of the river of the same name (Lamb and Bauer 2000; Matthee and Flemming 2002; Fritz et al. 2010; du Toit et al. 2012), where it forms a natural border between Namibia and South-Africa. It also corresponds to the southern limit of the desert area in this region (Hewitt 2004), as well as, together with the Bedford Gap (S5), the division between the Cape and other biogeographical subregions (Linder et al. 2012). Only reptiles exhibited consistent genetic discontinuities in this region (n = 3 studies). It is argued to restrict the gene flow among the northern and north-central lineages of the southern rock agama (*Agama atra*) (Matthee and Flemming 2002). However, for several taxa, the Orange River appears to only restrict population distribution to either its northern or southern banks, limiting the evidence of its role as primary cause of diversification. The first case can be observed for the common African four-striped mouse (*Rhabdomys pumilio*) (du Toit et al. 2012), whereas the second case was retrieved for the common rough gecko (*Pachydactylus barnardi*) (Lamb and Bauer 2000) and the sub-Saharan leopard tortoise (*Stigmochelys pardalis*) (Fritz et al. 2010). This river basin dates back to the Cretaceous, however, it has suffered considerable changes since then. During the Oligocene, its river-mouth was located 300-500 km south of its present location, and its modern course was only established towards the end of the Miocene (Goudie 2005). This dynamic evolution of the Orange river and its tributaries during the Cenozoic contributed to the contrasting patterns of genetic structure found in the area (Bauer 1999). Despite this, the timing of diversification between northern and southern lineages of the African horseshoe bat (*Rhinolophus damarensis*) is coincident with a late Miocene/Pliocene event, dated between 2.69 and 7.52 Mya (Jacobs et al. 2013). Both lineages are mainly restricted to each side of the Orange River, with the exception of one sampling site of the southern lineage located in central Namibia. The interplay between drainage dynamics, climatic oscillations, and habitat fluctuation/fragmentation would have promoted diversification of these lineages and their current genetic partitioning (Jacobs et al. 2013).

The Knersvlakte (S3): It is situated in the northern Western Cape province of South-Africa, being a 50 thousand-hectare arid plain, situated between the Bokkeveld and Kamiesberg Mountains. Being sparsely vegetated, mostly covered by pebbled terrain, it is considered a suitable habitat for rock-dwelling species. This genetic discontinuity revealed only as relevant in reptiles (n = 4 studies) among rupicolous or sparsely vegetation-dwelling species, including the angulate tortoise (*Chersina angulata*) (Daniels et al. 2007), the western rock skink (*Trachylepis sulcata*) (Portik, Bauer and Jackman 2011), the southern rock agama (*Agama atra*) (Matthee and Flemming 2002), and the common rough geckos of the genus *Pachydactylus* (Lamb and Bauer 2000). Nevertheless, the red rock hare (*Pronolagus rupestris*) also presents a genetic discontinuity in this region (Matthee and Robtnson 1996). Additionally, it was also found as a discontinuity for Western Cape endemic species, including the speckled padloper (*Chersobius signatus*) (Daniels et al. 2010) and the Cape rock elephant shrew (*Elephantulus edwardii*) (Smit, Robinson and Jansen van Vuuren 2007). The Knersvlakte may also represent a barrier for historical nuclear gene flow in mountain zebras (*Equus zebra*), although requiring further consideration (Moodley and Harley 2005). This region originated with the uplift of the Great Western Escarpment during the Miocene, around 18 Mya. However, divergence time estimates retrieved for the angulate tortoise (6.1-13.5 Mya) (Daniels et al. 2010), the southern rock agama (2.2-4.4 Mya) (Matthee and Flemming 2002) and the Cape rock elephant shrew (~2.4 Mya) (Smit et al. 2007), predate the completion time of the uplift. The origin of clades or incipient species in the southern rock agama (*Agama atra*), as well as in other reptile species, have instead been attributed to the climatic fluctuations during the past 2.5 Myr, leading to population fragmentation and isolation in mountainous refuges around the Knersvlakte region (Matthee and Flemming 2002). The specific conditions of the plain, therefore, could have been responsible for maintaining genetic distinctiveness between expanding populations.

Great Escarpment (S4): This is a 5 thousand km-long, semi-continuous mountain range system, forming a rim to the Southern African plateau (Clark, Barker and Mucina 2011). Although extending from north-western Angola, through Namibia, South-Africa, Zimbabwe, and Mozambique, the pattern retrieved in our analysis only includes two sections of the Great Escarpment, namely across the Namaqualand and the Western Karoo (as defined in Clark et al. 2011), in western South-Africa. The Great Escarpment was found to delimit relevant genetic discontinuities among reptiles (n = 2) and amphibians (n = 2), including the African clawed frog (*Xenopus laevis*) (Furman et al.

2015), the helmeted terrapin (*Pelomedusa galeata*) (Vamberger et al. 2018), and the African puff adder (*Bitis arietans*) (Barlow et al. 2013), with divergent populations occurring on both sides, as well as with the clicking stream frog (*Strongylopus grayii*) (Tolley, Braae and Cunningham 2010) distribution delimited to the southwestern portion of the Escarpment. The timing of divergence reported for this genetic discontinuity (280-950 kya; Barlow et al. 2013) dates to the Pleistocene, while its' current position was already defined by the end of the Cretaceous. Therefore, the genetic distribution of taxa across this discontinuity were probably shaped by both physiographic and climatic factors (Barlow et al. 2013; Vamberger et al. 2018).

Bedford Gap (S5): It comprehends an arid barrier situated between the Sundays and Great Fish river valleys, in South-Africa, which coincides with a contact zone of at least five diverse biomes: fynbos, succulent karoo, namakaroo, grassland, and savanna (Lawes 1990; Lawes et al. 2007; Willows-Munro and Matthee 2011). This discontinuity also delimits the transition between winter and summer rainfall zones, and although only retrieved as relevant for mammals (n = 5 studies), namely for the southern African shrew (*Myosorex varius*) (Willows-Munro and Matthee 2011), the southern African vlei rat (*Otomys irroratus*) (Engelbrecht et al. 2011), the Cape mole-rat (*Georychus capensis*) (Visser, Bennett and Jansen van Vuuren 2018), and the bontebuck (*Damaliscus pygargus*) (van der Walt, Nel and Hoelzel 2013), single studies on birds, amphibians, and invertebrates confirm the presence of a genetic discontinuity in this region. Among these are the sugarbirds of the genus *Promerops* (Haworth, Cunningham and Tjorve 2018), the clicking stream frog (*Strongylopus grayii*) (Tolley, Braae and Cunningham 2010), and the South African velvet worms of the genus *Peripatopsis* (Daniels et al. 2009). The rainfall seasonality does not seem to be the main driver of diversification, but rather a temporal barrier to gene flow (Tolley, Braae and Cunningham 2010). Marine transgressions during the early Pliocene (Siesser and Dingle 1981) coupled with multiple peaks of aridity during the Pleistocene (Stokes, Thomas and Washington 1997; DeMenocal 2004), would have promoted vicariance at the Bedford Gap, leading to the observed genetic patterns in this region. Divergence time estimates across several taxa coincide with Pliocene/Pleistocene vicariant events (Figure 2.5), namely for the southern African vlei rat (0.43-2.77 Mya) (Engelbrecht et al. 2011) and the southern African shrew (1.28-4.99 Mya) (Willows-Munro and Matthee 2011). Noteworthy is also the contact zone detected in this area, for the same species, resulting from the expansion of isolated populations during favourable climatic conditions in the Pleistocene (Engelbrecht et al. 2011; Willows-Munro and Matthee 2011).

Within Southern Africa, our analysis did not retrieve a putative refugium to the southwest, described in previous studies and covering an area represented by Angola and coastal Namibia (Lorenzen et al. 2012; Bertola et al. 2016). Two main reasons may explain this, including: i) the area being poorly sampled (Figure 2.3 and Figure 2.4), and ii) the area being very wide, leading to methodological constrains in detecting it, since patterns representing genetic discontinuities for different species may have been displaced apart, and not considered for the cumulative analysis. For example, the restricted distribution of the Giant sable antelope (*Hippotragus niger variani*) in Angola, was represented in our analysis by a small discontinuity in this area (Pitra et al. 2002; Jansen van Vuuren et al. 2010); the mountain zebra (*Equus zebra*) (Pedersen et al. 2018) and the greater kudu (*Tragelaphus strepsiceros*) (Nersting and Arctander 2001), have restricted Namibian populations; and an impala (*Aepyceros melampus*) lineage is located between the Angolan and the Namibian border (Nersting and Arctander 2001; Grobler et al. 2017). Despite the lack of cumulative patterns to define the Southwest region, we underlie its importance as a possible refugium, geographically separated from other southern populations by the extremely arid Kalahari Desert (Nersting and Arctander 2001; Lorenzen et al. 2012).

Olson et al. (2001) also presented subdivisions into ecoregions among West, East, and South of Africa, whose limits can be linked to some of the genetic discontinuities found in this review [e.g. Dahomey-Volta (W2) and Dahomey-Niger (W3) complexes, Eastern climatic break (E2), Zambezi-Kafue complex (S2) and Great Escarpment (S4)].

2.4 Future prospects and challenges

This review highlights the importance of analysing genetic discontinuities that have shaped current species distribution in a region, considering both its climatic and geomorphological history. Nonetheless, it is important to acknowledge that the distribution of diversity across sub-Saharan Africa is not solely influenced by these factors. Human disturbance is proven to have a tremendous impact on species' distribution across African forests and savanna biomes (Hannah, Carr and Lankerani 1995; Malhi et al. 2013). In recent decades, increased human population growth has intensified this disturbance through anthropogenic activities, such as urbanization, commercial logging, mining, and conversion of forests into farmland (Western, Groom and Worden 2009; Laurance, Sayer and Cassman 2014; Ola and Benjamin 2019). Savanna vegetation, in particular, has been identified as highly vulnerable to the effects of climate change (Midgley and Bond 2015). Therefore, future phylogeographic assessments across sub-Saharan Africa should consider the combination of both natural

and anthropogenic factors influencing the distribution of diversity for this region. It is also important to emphasize that several taxonomic groups might be more vulnerable to specific set of factors (e.g amphibians), and research bias towards other group, may enhance this tendency.

2.5 References

Aghová, T., Šumbera, R., Piálek, L., Mikula, O., McDonough, M. M., Lavrenchenko, L. A., meheretu, Y., Mbau, J. S. and Bryja, J., 2017. Multilocus phylogeny of East African gerbils (Rodentia, *Gerbilliscus*) illuminates the history of the Somali-Masai savanna. *Journal of Biogeography*, 44: 2295–2307.

Aldenderfer, M., 2006. Modelling plateau peoples: the early human use of the world's high plateaux. *World Archaeology*, 38: 357–370.

Alfaro, M. E., Zoller, S. and Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, 20: 255–266.

Allal, F., Sanou, H., Millet, L., Vaillant, A., Camus-Kulandaivelu, L., Logossa, Z. A., Lefèvre, F. and Bouvet, J. M., 2011. Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. *Heredity*, 107: 174–186.

Allio, R., Donega, S., Galtier, N. and Nabholz, B., 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular. *Molecular Biology and Evolution*, 34: 2762–2772.

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Anhuf, D., 2000. Vegetation history and climate changes in Africa north and south of the equator (10 S to 10 N) during the last glacial maximum, pp. 225-248 in *Southern hemisphere paleo-and neoclimates*. Springer (P. Smolka and W. Volkheimer, Eds.), Berlin.

Arbogast, B. S. and Kenagy, G. J., 2001. Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, 28: 819–825.

Arctander, P., Johansen, C. and Coutellec-Vreto, M.-A., 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution*, 16: 1724–1739.

Avise, J., Arnold, J., Ball, R. and Bermingham, E., 1987. Intraspecific phylogeography:

the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18: 489–522.

Avise, J. C., 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, 7: 371–379.

Avise, J. C., 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachusetts.

Avise, J. C., 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36: 3–15.

Badgley, C., 2010. Tectonics, topography, and mammalian diversity. *Ecography*, 33: 220–231.

Bagley, J. C. and Johnson, J. B., 2014. Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biological Reviews*, 89: 767–790.

Ballard, J. W. O. and Whitlock, M. C., 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, 13: 729-744.

Barlow, A., Baker, K., Hendry, C. R., Peppin, L., Phelps, T., Tolley, K. A., Wüster, C. E. and Wüster, W., 2013. Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology*, 22: 1134–1157.

Barlow, A., Wüster, W., Kelly, C. M., Branch, W. R., Phelps, T. and Tolley, K. A., 2019. Ancient habitat shifts and organismal diversification are decoupled in the African viper genus *Bitis* (Serpentes: Viperidae). *Journal of Biogeography*, 46: 1234–1248.

Barratt, C. D., Bwong, B. A., Jehle, R., Liedtke, H. C., Nagel, P., Onstein, R. E., Portik, D. M., Streicher, J. W. and Loader, S. P., 2018. Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Molecular Ecology*, 27: 4289–4308.

Bartáková, V., Reichard, M., Blažek, R., Polačik, M. and Bryja, J., 2015. Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna. *Journal of Biogeography*, 42: 1832–1844.

Bauer, A. M., 1999. Evolutionary scenarios in the *Pachydactylus* group geckos of southern Africa: new hypotheses. *African Journal of Herpetology*, 48: 53–62.

Beheregaray, L. B., 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17: 3754–3774.

Belay, G. and Mori, A., 2006. Intraspecific phylogeographic mitochondrial DNA (D-loop) variation of Gelada baboon, *Theropithecus gelada*, in Ethiopia. *Biochemical Systematics and Ecology*, 34: 554–561.

Bermingham, E. and Moritz, C., 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology*, 7: 367–369.

Bertola, L. D., Jongbloed, H., Van Der Gaag, K. J., de Kniijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., et al., 2016. Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the Lion (*Panthera leo*). *Scientific Reports*, 6: 30807.

Bezault, E., Balaresque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J. F. and Rognon, X., 2011. Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics*, 12: 1–16.

Bibi, F. and Kiessling, W., 2015. Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences*, 112: 10623–10628.

Bittencourt-Silva, G. B., Lawson, L. P., Tolley, K. A., Portik, D. M., Barratt, C. D., Nagel, P. and Loader, S. P., 2017. Impact of species delimitation and sampling on niche models and phylogeographical inference: a case study of the East African reed frog *Hyperolius substriatus* Ahl, 1931. *Molecular Phylogenetics and Evolution*, 114: 261–270.

Bobe, R. and Eck, G., 2001. Responses of African bovids to Pliocene climate change. *Paleobiology*, 27: 1–47.

Bobe, R., Behrensmeyer, A. K. and Chapman, R. E., 2002. Faunal change, environmental variability and late Pliocene hominin evolution. *Journal of Human Evolution*, 42: 475–497.

Bobe, R., 2006. The evolution of arid ecosystems in eastern Africa. *Journal of Arid Environments*, 66: 564–584.

Bohoussou, K. H., Cornette, R., Akpatou, B., Colyn, M., Kerbis Peterhans, J., Kennis, J., et al., 2015. The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. *Journal of Biogeography*, 42: 2049–2061.

Booth, A. H., 1958. The Niger, the Volta and the Dahomey Gap as geographic barriers. *Evolution*, 12:48–62.

Bowie, R. C., Fjeldså, J., Hackett, S. J. and Crowe, T. M., 2004. Molecular evolution in space and through time: mtDNA phylogeography of the Olive Sunbird (*Nectarinia olivacea/obscura*) throughout continental Africa. *Molecular Phylogenetics and Evolution*, 33: 56–74.

Bowman, D.M.J.S., 2000. Tropical Rain Forests. *Progress in Physical Geography*, 24: 103–109.

Boyer, A.G. and Jetz W., 2012. Conservation biology, pp. 271-279 in *Metabolic ecology: a scaling approach*. Wiley-Blackwell Publications (R. M. Sibly, J. H. Brown and A. Kodric-Brown, Eds.), Oxford.

Brouat, C., Tatard, C., Bâ, K., Cosson, J. F., Dobigny, G., Fichet-Calvet, E., et al., 2009. Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): a case study for Sahelian species in West Africa. *Journal of Biogeography*, 36: 2237–2250.

Brown, J. H. and Maurer, B. A., 1987. Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *The American Naturalist*, 130: 1–17.

Bryja, J., Mikula, O., Šumbera, R., Meheretu, Y., Aghová, T., Lavrenchenko, L. A., et al., 2014. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology*, 14: 1–20.

Bryja, J., Konvičková, H., Bryjová, A., Mikula, O., Makundi, R. and Chitaukali, W. N., 2018. Differentiation underground: range-wide multilocus genetic structure of the silvery mole-rat does not support current taxonomy based on mitochondrial sequences. *Mammalian Biology*, 93: 82–92.

Burgess, N. D., Clarke, G. P. and Rodgers, W. A., 1998. Coastal forests of eastern Africa: status, endemism patterns and their potential causes. *Biological Journal of the Linnean Society*, 64: 337–367.

Burgess, N., Butynski, T. M., Cordeiro, N. J., Doggart, N. H., Fjeldså, J., Howell, K. M., et al., 2007. The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134: 209–231.

Burress, E. D., 2015. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia*, 748: 7–27.

Burrough, S., Thomas, D. and Bailey, R., 2009. Mega-Lake in the Kalahari: a Late Pleistocene record of the Palaeolake Makgadikgadi system. *Quaternary Science Reviews*, 28: 1392–1411.

Carlen, E. J., Rathbun, G. B., Olson, L. E., Sabuni, C. A., Stanley, W. T. and Dumbacher, J. P., 2017. Reconstructing the molecular phylogeny of giant sengis (Macroscelidea; Macroscelididae; *Rhynchocyon*). *Molecular Phylogenetics and Evolution*, 113: 150–160.

Castiglia, R., Solano, E., Makundi, R. H., Hulselmans, J., Verheyen, E. and Colangelo, P., 2012. Rapid chromosomal evolution in the mesic four-striped grass rat *Rhabdomys dilectus* (Rodentia, Muridae) revealed by mtDNA phylogeographic analysis. *Journal of Zoological Systematics and Evolutionary Research*, 50: 165–172.

Chala, D., Zimmermann, N. E., Brochmann, C. and Bakkestuen, V., 2017. Migration corridors for alpine plants among the 'sky islands' of eastern Africa: do they, or did they exist? *Alpine Botany*, 127: 133–144.

Charruau, P., Fernandes, C., Orozco-Terwengel, P., Peters, J., Hunter, L., Ziaie, H., Jourabchian, A., Jowkar, H., et al., 2011. Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. *Molecular Ecology,* 20: 706-724.

Chen, L. Y., Muchuku, J. K., Yan, X., Hu, G. W. and Wang, Q. F., 2015. Phylogeography of *Haplocarpha rueppelii* (Asteraceae) suggests a potential geographic barrier for plant dispersal and gene flow in East Africa. *Science Bulletin*, 60: 1184–1192.

Chorowicz, J., 2005. The East African rift system. *Journal of African Earth Sciences*, 43: 379–410.

Clark, V. R., Barker, N. P. and Mucina, L., 2011. The Great Escarpment of southern Africa: a new frontier for biodiversity exploration. *Biodiversity and Conservation*, 20: 2543–2561.

Coals, P. G. and Rathbun, G. B., 2013. The taxonomic status of giant sengis (genus *Rhynchocyon*) in Mozambique. *Journal of East African Natural History*, 101: 241–250.

Colangelo, P., Verheyen, E., Leirs, H., Tatard, C., Denys, C., Dobigny, G., Duplantier, J-M., Brouat, C., Granjon, L. and Lecompte, E., 2013. A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society*, 108: 901–916.

Colyn, M., Hulselmans, J., Sonet, G., Oude, P., De Winter, J., Natta, A., Nagy, Z. T. and Verheyen, E., 2010. Discovery of a new duiker species (Bovidae: Cephalophinae) from the Dahomey Gap, West Africa. *Zootaxa*, 2637: 1–30.

Couvreur, T. L., Dauby, G., Blach-Overgaard, A., Deblauwe, V., Dessein, S., Droissart, V., et al., 2020. Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. *Biological Reviews*, 96: 16–51.

Daïnou, K., Mahy, G., Duminil, J., Dick, C. W., Doucet, J. L., Donkpégan, A. S. L., Pluijgers, M., Sinsin, B., Lejeune, P. and Hardy, O. J., 2014. Speciation slowing down in widespread and long-living tree taxa: insights from the tropical timber tree genus *Milicia* (Moraceae). *Heredity*, 113: 74–85.

Daniels, S. R., Hofmeyr, M. D., Henen, B. T. and Crandall, K. A., 2007. Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. *Molecular Phylogenetics and Evolution*, 45: 915–926.

Daniels, S. R., Picker, M. D., Cowlin, R. M. and Hamer, M. L., 2009. Unravelling evolutionary lineages among South African velvet worms (Onychophora: *Peripatopsis*) provides evidence for widespread cryptic speciation. *Biological Journal of the Linnean Society*, 97: 200–216.

Daniels, S. R., Hofmeyr, M. D., Henen, B. T. and Baard, E. H. W., 2010. Systematics and phylogeography of a threatened tortoise, the speckled padloper. *Animal Conservation*, 13: 237–246.

Dehghani, R., Wanntorp, L., Pagani, P., Källersjö, M., Werdelin, L. and Veron, G., 2008. Phylogeography of the white-tailed mongoose (Herpestidae, Carnivora, Mammalia) based on partial sequences of the mtDNA control region. *Journal of Zoology*, 276: 385–393.

DeJong, R. J., Morgan, J. A. T., Wilson, W. D., Al-Jaser, M. H., Appleton, C. C., Coulibaly, G., et al., 2003. Phylogeography of *Biomphalaria glabrata* and *B. pfeifferi*, important intermediate hosts of *Schistosoma mansoni* in the New and Old World tropics. *Molecular Ecology*, 12: 3041–3056.

DeMenocal, P. B., 1995. Plio-Pleistocene African climate. *Science*, 270: 53–59.

DeMenocal, P. B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, 220: 3–24.

Demenou, B. B., Piñeiro, R. and Hardy, O. J., 2016. Origin and history of the Dahomey Gap separating West and Central African rain forests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. *Journal of Biogeography*, 43: 1020–1031.

Demenou, B. B., Doucet, J. L. and Hardy, O. J., 2018. History of the fragmentation of the African rain forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*. *Heredity*, 120: 547–561.

Denys, C., Chorowicz, J. and Tiercelin, J. J., 1986. Tectonic and environmental control on rodent diversity in the Plio-Pleistocene sediments of the African Rift System. *Geological Society, London, Special Publications*, 25: 363–372.

de Wasseige, C., Flynn, J., Louppe, D., Hiol Hiol, F. and Mayaux P., 2014. *The Forests of the Congo Basin – State of the Forest 2013*. Weyrich, Neufchâteau.

Dobigny, G., Tatard, C., Gauthier, P., Ba, K., Duplantier, J. M., Granjon, L. and Kergoat, G. J., 2013. Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan open habitats Pleistocene history. Plos One, 8: e77815.

Doucouré, C. and de Wit, M., 2003. Old inherited origin for the present near-bimodal topography of Africa. *Journal of African Earth Sciences*, 36: 371–388.

Duminil, J., Brown, R. P., Ewédjè, E. E. B., Mardulyn, P., Doucet, J. L. and Hardy, O. J.,

2013. Large-scale pattern of genetic differentiation within African rainforest trees: insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp (Fabaceae). *BMC Evolutionary Biology*, 13: 1–13.

Dupont, L. M., Jahns, S., Marret, F. and Ning, S., 2000. Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 155: 95–122.

Dupont, L., 2011. Orbital scale vegetation change in Africa. *Quaternary Science Reviews*, 30: 3589–3602.

Durand, A., 1982. Oscillations of Lake Chad over the past 50,000 years: New data and new hypothesis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 39: 37–53.

du Toit, N., Jansen van Vuuren, B., Matthee, S. and Matthee, C. A., 2012. Biome specificity of distinct genetic lineages within the four-striped mouse *Rhabdomys pumilio* (Rodentia: Muridae) from southern Africa with implications for taxonomy. *Molecular Phylogenetics and Evolution*, 65: 75–86.

Eaton, M. J., Martin, A., Thorbjarnarson, J. and Amato, G., 2009. Species-level diversification of African dwarf crocodiles (Genus *Osteolaemus*): a geographic and phylogenetic perspective. *Molecular Phylogenetics and Evolution*, 50: 496–506.

Engelbrecht, A., Taylor, P. J., Daniels, S. R. and Rambau, R. V., 2011. Cryptic speciation in the southern African vlei rat *Otomys irroratus* complex: evidence derived from mitochondrial cyt b and niche modelling. *Biological Journal of the Linnean Society*, 104: 192–206.

Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J. and Cannatella, D. C., 2004. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution*, 33: 197–213.

Feakins, S. J., DeMenocal, P. B. and Eglinton, T. I., 2005. Biomarker records of late Neogene changes in northeast African vegetation. *Geology*, 33: 977–980.

Fisher-Reid, M. C. and Wiens, J. J., 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. *BMC Evolutionary Biology*, 11: 1–20.

Fjeldså, J., Burgess, N. D., Blyth, S. and De Klerk, H. M., 2004. Where are the major gaps in the reserve network for Africa's mammals? *Oryx*, 38: 17–25.

Flagstad, Ø., Syversten, P. O., Stenseth, N. C. and Jakobsen, K. S., 2001. Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climatic variation. *Proceedings of the Royal Society of London B: Biological Sciences*, 268: 667–677.

Foster, A., Ebinger, C., Mbede, E. and Rex, D., 1997. Tectonic development of the northern Tanzanian sector of the East African rift system. *Journal of the Geological Society*, 154: 689–700.

Freilich, X., Anadón, J. D., Bukala, J., Calderon, O., Chakraborty, R. and Boissinot, S., 2016. Comparative Phylogeography of Ethiopian anurans: impact of the Great Rift Valley and Pleistocene climate change. *BMC Evolutionary Biology*, 16: 1–19.

Fritz, U., Daniels, S. R., Hofmeyr, M. D., González, J., Barrio-Amorós, C. L., Široký, P., Hundsdörfer, A. K. and Stuckas, H., 2010. Mitochondrial phylogeography and subspecies of the wide-ranging sub-Saharan leopard tortoise *Stigmochelys pardalis* (Testudines: Testudinidae) – a case study for the pitfalls of pseudogenes and GenBank sequences. *Journal of Zoological Systematics and Evolutionary Research*, 48: 348–359.

Fuchs, J., Fjeldså, J. and Bowie, R. C., 2017. Diversification across major biogeographic breaks in the African Shining/Square-tailed Drongos complex (Passeriformes: Dicruridae). *Zoologica Scripta*, 46: 27–41.

Furman, B. L., Bewick, A. J., Harrison, T. L., Greenbaum, E., Gvoždík, V., Kusamba, C. and Evans, B. J., 2015. Pan-African phylogeography of a model organism, the African clawed frog '*Xenopus laevis'*. *Molecular Ecology*, 24: 909–925.

Gaston, K. J., 2000. Global patterns in biodiversity. *Nature*, 405: 220–227.

Gaubert, P., Njiokou, F., Ngua, G., Afiademanyo, K., Dufour, S., Malekani, J., et al., 2016. Phylogeography of the heavily poached African common pangolin (Pholidota, *Manis tricuspis*) reveals six cryptic lineages as traceable signatures of Pleistocene diversification. *Molecular Ecology*, 25: 5975–5993.

Ghienne, J. F., Schuster, M., Bernard, A., Duringer, P. and Brunet, M., 2002. The Holocene giant Lake Chad revealed by digital elevation models. *Quaternary International*, 87: 81–85.

Gichira, A. W., Li, Z. Z., Saina, J. K., Hu, G. W., Gituru, R. W., Wang, Q. F. and Chen, J. M., 2017). Demographic history and population genetic structure of *Hagenia abyssinica* (Rosaceae), a tropical tree endemic to the Ethiopian highlands and eastern African mountains. *Tree Genetics & Genomes*, 13: 1–11.

Gizaw, A., Kebede, M., Nemomissa, S., Ehrich, D., Bekele, B., Mirré, V., Popp, M. and Brochmann, C., 2013. Phylogeography of the heathers *Erica arborea* and *E. trimera* in the afro-alpine 'sky islands' inferred from AFLPs and plastid DNA sequences. *Flora – Morphology, Distribution, Functional Ecology of Plants*, 208: 453–463.

Gizaw, A., Wondimu, T., Mugizi, T. F., Masao, C. A., Abdi, A. A., Popp, M., Ehrich, D., Nemomissa, S. and Brochmann, C., 2016. Vicariance, dispersal, and hybridization

in a naturally fragmented system: the afro-alpine endemics *Carex monostachya* and *C. runssoroensis* (Cyperaceae). *Alpine Botany*, 126: 59–71.

Goodier, S. A., Cotterill, F. P., O'Ryan, C., Skelton, P. H. and de Wit, M. J., 2011. Cryptic diversity of African tigerfish (Genus *Hydrocynus*) reveals palaeogeographic signatures of linked Neogene geotectonic events. *Plos One*, 6: e28775.

Gottelli, D., Marino, J., Sillero-Zubiri, C. and Funk, S. M., 2004. The effect of the last glacial age on speciation and population genetic structure of the endangered Ethiopian wolf (*Canis simensis*). *Molecular Ecology*, 13: 2275–2286.

Goudie, A. S., 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67: 437–456.

Grace, J., José, J. S., Meir, P., Miranda, H. S. and Montes, R. A., 2006. Productivity and carbon fluxes of tropical savannas. *Journal of Biogeography*, 33: 387–400.

Griffin, D. L., 2006. The late Neogene Sahabi rivers of the Sahara and their climatic and environmental implications for the Chad Basin. *Journal of the Geological Society*, 163: 905–921.

Griffiths, C. J., 1993. The geological evolution of East Africa, pp. 9-21 in *Biogeography and ecology of the rain forests of eastern Africa*. Cambridge University Press (J. Lovett and K. Wasser, Eds.), Cambridge.

Grobler, J. P., Hayter, K. N., Labuschagne, C., Nel, E. J. and Coetzer, W. G., 2017. The genetic status of naturally occurring black-nosed impala from northern South Africa. *Mammalian Biology*, 82: 27–33.

Grove, A. T., 1986. Geomorphology of the African Rift system. *Geological Society, London, Special Publications*, 25: 9–16.

Guiraud, R. and Maurin, J. C., 1992. Early Cretaceous rifts of Western and Central Africa: an overview. *Tectonophysics*, 213: 153–168.

Hamilton, A. A. and Taylor, D., 1991. History of climate and forests in tropical Africa during the last 8 million years, pp. 65-78 in *Tropical forests and climate*. Springer (N. Myers, Eds.), Dordrecht.

Hannah, L., Carr, J. L. and Lankerani, A., 1995. Human disturbance and natural habitat: a biome level analysis of a global data set. *Biodiversity & Conservation*, 4: 128–155.

Hardy, O. J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J., et al., 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, 345: 284–296.

Hassanin, A., Khouider, S., Gembu, G. C., Goodman, S. M., Kadjo, B., Nesi, N., Pourrut, X., Nakouné, E. and Bonillo, C., 2015. The comparative phylogeography of fruit bats of the tribe Scotonycterini (Chiroptera, Pteropodidae) reveals cryptic species

diversity related to African Pleistocene forest refugia. *Comptes Rendus Biologies*, 338: 197–211.

Hassanin, A., Nesi, N., Marin, J., Kadjo, B., Pourrut, X., Leroy, É., et al., 2016. Comparative phylogeography of African fruit bats (Chiroptera, Pteropodidae) provide new insights into the outbreak of Ebola virus disease in West Africa, 2014–2016. *Comptes Rendus Biologies*, 339: 517–528.

Haus, T., Akom, E., Agwanda, B., Hofreiter, M., Roos, C. and Zinner, D., 2013. Mitochondrial diversity and distribution of African green monkeys (*Chlorocebus* Gray, 1870). *American Journal of Primatology*, 75: 350–360.

Haworth, E. S., Cunningham, M. J. and Tjorve, K. M. C., 2018. Population diversity and relatedness in Sugarbirds (Promeropidae: *Promerops* spp.). *PeerJ*, 6: e5000.

Hewitt, G., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58: 247–276.

Hewitt, G., 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68: 87–112.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405: 907–913.

Hewitt, G., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359: 183–195.

Hickerson, M. J., Dolman, G. and Moritz, C., 2006. Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology*, 15: 209–223.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F. and Yoder, A. D., 2010. Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, 54: 291–301.

Hildebrand, E. A., Brandt, S. A., Friis, I. and Demissew, S., 2019. Paleoenvironmental reconstructions for the Horn of Africa: Interdisciplinary perspectives on strategy and significance, pp. 187-210 in *Trees, grasses and crops. People and plants in sub-Saharan Africa and beyond*. Verlag Dr. Rudolf Habelt GmbH (B. Eichhorn and A. Höhn, Eds.), Bonn.

Huntley, J. W. and Voelker, G., 2016. Cryptic diversity in Afro-tropical lowland forests: the systematics and biogeography of the avian genus *Bleda*. *Molecular Phylogenetics and Evolution*, 99: 297–308.

Huntley, J. W., Harvey, J. A., Pavia, M., Boano, G. and Voelker, G., 2018. The systematics and biogeography of the Bearded Greenbuls (Aves: *Criniger*) reveals the impact of Plio-Pleistocene forest fragmentation on Afro-tropical avian diversity.

Zoological Journal of the Linnean Society, 183: 672–686.

Jacobs, B. F., 2004. Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359: 1573–1583.

Jacobs, D. S., Babiker, H., Bastian, A., Kearney, T., van Eeden, R. and Bishop, J. M., 2013. Phenotypic convergence in genetically distinct lineages of a *Rhinolophus* species complex (Mammalia, Chiroptera). *Plos One*, 8: e82614.

Jacquet, F., Denys, C., Verheyen, E., Bryja, J., Hutterer, R., Peterhans, J. C. K., et al., 2015. Phylogeography and evolutionary history of the *Crocidura olivieri* complex (Mammalia, Soricomorpha): from a forest origin to broad ecological expansion across Africa. *BMC Evolutionary Biology*, 15: 1–15.

Jansen van Vuuren, B., Robinson, T. J., Vaz Pinto, P., Estes, R. and Matthee, C. A., 2010. Western Zambian sable: are they a geographic extension of the giant sable antelope? *South African Journal of Wildlife Research*, 40: 35–42.

Jarvis, J.U.M., 2013. Genus Tachyoryctes - Root-Rats, pp. 148–52, in *Mammals of Africa III: Rodents, Hares and Rabbits*. Bloomsbury Publishing (D. Happold, J. Kingdon and M. Happold, Eds.), London

Jess, S., Koehn, D., Fox, M., Enkelmann, E., Sachau, T. and Aanyu, K., 2020. Paleogene initiation of the Western Branch of the East African Rift: The uplift history of the Rwenzori Mountains, Western Uganda. *Earth and Planetary Science Letters*, 552: 116593.

Jongsma, G. F., Barej, M. F., Barratt, C. D., Burger, M., Conradie, W., Ernst, R., et al., 2018. Diversity and biogeography of frogs in the genus *Amnirana* (Anura: Ranidae) across sub-Saharan Africa. *Molecular Phylogenetics and Evolution*, 120: 274–285.

Kadu, C. A., Konrad, H., Schueler, S., Muluvi, G. M., Eyog-Matig, O., Muchugi, A., et al., 2013. Divergent pattern of nuclear genetic diversity across the range of the Afromontane *Prunus africana* mirrors variable climate of African highlands. *Annals of Botany*, 111: 47–60.

Kaya, F., Bibi, F., Žliobaitė, I., Eronen, J. T., Hui, T. and Fortelius, M., 2018. The rise and fall of the Old World savannah fauna and the origins of the African savannah biome. *Nature Ecology & Evolution*, 2: 241–246.

Kebede, M., Ehrich, D., Taberlet, P., Nemomissa, S. and Brochmann, C., 2007. Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology*, 16: 1233–1243.

Kidd, D. M. and Ritchie, M. G., 2006. Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, 33: 1851–1865.

Kide, N. G., Dunz, A., Agnèse, J. F., Dilyte, J., Pariselle, A., Carneiro, C., et al., 2016. Cichlids of the Banc d'Arguin National Park, Mauritania: insight into the diversity of the genus *Coptodon*. *Journal of Fish Biology*, 88: 1369–1393.

Kim, Y. and Eltahir, E. A., 2004. Role of topography in facilitating coexistence of trees and grasses within savannas. *Water Resources Research*, 40.

Kissling, W. D., Eiserhardt, W. L., Baker, W. J., Borchsenius, F., Couvreur, T. L., Balslev, H. and Svenning, J. C., 2012. Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. *Proceedings of the National Academy of Sciences*, 109: 7379–7384.

Koffi, K. G., Hardy, O. J., Doumenge, C., Cruaud, C. and Heuertz, M., 2011. Diversity gradients and phylogeographic patterns in *Santiria trimera* (Burseraceae), a widespread African tree typical of mature rainforests. *American Journal of Botany*, 98: 254–264.

Lamb, T. and Bauer, A. M., 2000. Relationships of the *Pachydactylus rugosus* group of geckos (Reptilia: Squamata: Gekkonidae). *African Zoology*, 35: 55–67.

Lamb, J. M., Ralph, T., Goodman, S. M., Bogdanowicz, W., Fahr, J., Gajewska, M., Bates, P. J. J., Eger, J., Petr, B. and Taylor, P. J., 2008. Phylogeography and predicted distribution of African-Arabian and Malagasy populations of giant mastiff bats, *Otomops* spp.(Chiroptera: Molossidae). *Acta Chiropterologica*, 10: 21–40.

Lancaster, N., 1989. Late Quaternary paleoenvironments in the southwestern Kalahari. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 70: 367–376.

Laurance, W. F., Sayer, J. and Cassman, K. G., 2014. Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, 29: 107–116.

Lawes, M. J., 1990. The distribution of the samango monkey (*Cercopithecus mitis erythrarchus* Peters, 1852 and *Cercopithecus mitis labiatus I.* Geoffroy, 1843) and forest history in southern Africa. *Journal of Biogeography*, 17: 669–680.

Lawes, M. J., Eeley, H. A., Findlay, N. J. and Forbes, D., 2007. Resilient forest faunal communities in South Africa: a legacy of palaeoclimatic change and extinction filtering? *Journal of Biogeography*, 34: 1246–1264.

Lawson, L. P., 2013. Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the African spotted reed frog. *Molecular Ecology*, 22: 1947–1960.

Leaché, A. D. and Fujita, M. K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences*, 277: 3071–3077.

Leite, R. N. and Rogers, D. S., 2013. Revisiting Amazonian phylogeography: insights into diversification hypotheses and novel perspectives. *Organisms Diversity & Evolution*, 13: 639–664.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J. and Rahbek, C., 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39: 1189–1205.

Linder, H. P., 2017. East African Cenozoic vegetation history. *Evolutionary Anthropology: Issues, News, and Reviews*, 26: 300–312.

Lobon, I., Tucci, S., de Manuel, M., Ghirotto, S., Benazzo, A., Prado-Martinez, J., et al., 2016. Demographic history of the genus *Pan* inferred from whole mitochondrial genome reconstructions. *Genome Biology and Evolution*, 8: 2020–2030.

Lompo, D., Vinceti, B., Konrad, H., Gaisberger, H. and Geburek, T., 2018. Phylogeography of African locust bean (*Parkia biglobosa*) reveals genetic divergence and spatially structured populations in West and Central Africa. *Journal of Heredity*, 109: 811–824.

Lorenzen, E. D., Heller, R. and Siegismund, H. R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21: 3656–3670.

Lovett, J. C., 1993. Climatic history and forest distribution in eastern Africa, pp. 23-29 in *Biogeography and ecology of the rain forests of eastern Africa*. Cambridge University Press (J. Lovett and K. Wasser, Eds.), Cambridge.

Lyons, R. P., Scholz, C. A., Cohen, A. S., King, J. W., Brown, E. T., Ivory, S. J., Johnson, T. C., Deino, A. L., et al., 2015. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proceedings of the National Academy of Sciences*, 112: 15568–15573.

Mairal, M., Sanmartín, I. and Pellissier, L., 2017. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. *Journal of Biogeography*, 44: 911–923.

Mairal, M., Sanmartín, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J. J. and Alarcón, M., 2017. Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports*, 7: 1–13.

Maley, J., 1996. The African rain forest – main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh. Section B. Biological Sciences*, 104: 31–73.

Malhi, Y., Adu-Bredu, S., Asare, R. A., Lewis, S. L. and Mayaux, P., 2013. African rainforests: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368: 20120312.

Manthey, J. D., Reyes-Velasco, J., Freilich, X. and Boissinot, S., 2017. Diversification in

a biodiversity hotspot: genomic variation in the river frog *Amietia nutti* across the Ethiopian Highlands. *Biological Journal of the Linnean Society*, 122: 801–813.

Marks, B. D., 2010. Are lowland rainforests really evolutionary museums? Phylogeography of the green hylia (*Hylia prasina*) in the Afrotropics. *Molecular Phylogenetics and Evolution*, 55: 178–184.

Marzoli, A., Piccirillo, E. M., Renne, P. R., Bellieni, G., Iacumin, M., Nyobe, J. B. and Tongwa, A. T., 2000. The Cameroon Volcanic Line revisited: petrogenesis of continental basaltic magmas from lithospheric and asthenospheric mantle sources. *Journal of Petrology*, 41: 87–109.

Maslin, M. A. and Christensen, B., 2007. Tectonics, orbital forcing, global climate change, and human evolution in Africa: introduction to the African paleoclimate special volume. *Journal of Human Evolution*, 53: 443–464.

Maslin, M. A., Pancost, R. D., Wilson, K. E., Lewis, J. and Trauth, M. H., 2012. Three and half million year history of moisture availability of South West Africa: evidence from ODP site 1085 biomarker records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 317: 41–47.

Matthee, C. A. and Robinson, T. J., 1996. Mitochondrial DNA differentiation among geographical populations of *Pronolagus rupestris*, Smith's red rock rabbit (Mammalia: Lagomorpha). *Heredity*, 76: 514–523.

Matthee, C. A. and Robinson, T. J., 1997. Mitochondrial DNA phylogeography and comparative cytogenetics of the springhare, *Pedetes capensis* (Mammalia: Rodentia). *Journal of Mammalian Evolution*, 4: 53–73.

Matthee, C. A. and Flemming, A. F., 2002. Population fragmentation in the southern rock agama, *Agama atra*: more evidence for vicariance in Southern Africa. *Molecular Ecology*, 11: 465–471.

Mayaux, P., Pekel, J. F., Desclée, B., Donnay, F., Lupi, A., Achard, F., et al., 2013. State and evolution of the African rainforests between 1990 and 2010. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368: 20120300.

Mazoch, V., Mikula, O., Bryja, J., Konvičková, H., Russo, I. R., Verheyen, E. and Šumbera, R., 2018. Phylogeography of a widespread sub-Saharan murid rodent *Aethomys chrysophilus*: the role of geographic barriers and paleoclimate in the Zambezian bioregion. *Mammalia*, 82: 373–387.

McDonough, M. M., Šumbera, R., Mazoch, V., Ferguson, A. W., Phillips, C. D. and Bryja, J., 2015. Multilocus phylogeography of a widespread savanna–woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology*, 24: 5248–5266.

Measey, G. J., Galbusera, P., Breyne, P. and Matthysen, E., 2007. Gene flow in a direct-

developing, leaf litter frog between isolated mountains in the Taita Hills, Kenya. *Conservation Genetics*, 8: 1177–1188.

Menegon, M., Loader, S. P., Marsden, S. J., Branch, W. R., Davenport, T. R. and Ursenbacher, S., 2014. The genus *Atheris* (Serpentes: Viperidae) in East Africa: phylogeny and the role of rifting and climate in shaping the current pattern of species diversity. *Molecular Phylogenetics and Evolution*, 79: 12–22.

Midgley, G. F. and Bond, W. J., 2015. Future of African terrestrial biodiversity and ecosystems under anthropogenic climate change. *Nature Climate Change*, 5: 823–829.

Migliore, J., Kaymak, E., Mariac, C., Couvreur, T. L., Lissambou, B. J., Piñeiro, R. and Hardy, O. J., 2019. Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from *Greenwayodendron* (Annonaceae) phylogenomics. *Journal of Biogeography*, 46: 212–223.

Mikula, O., Šumbera, R., Aghova, T., Mbau, J. S., Katakweba, A. S., Sabuni, C. A. and Bryja, J., 2016. Evolutionary history and species diversity of African pouched mice (Rodentia: Nesomyidae: *Saccostomus*). *Zoologica Scripta*, 45: 595–617.

Miller, J. M., Hallager, S., Monfort, S. L., Newby, J., Bishop, K., Tidmus, S. A., Black, P., Houston, B., et al., 2011. Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). *Conservation Genetics*, 12: 423–431.

Missoup, A. D., Nicolas, V., Wendelen, W., Bilong, C. F. B., Cruaud, C. and Denys, C., 2009. First record of *Hylomyscus walterverheyeni* (Rodentia: Muridae) on the north-western side of the Sanaga River (western Cameroon). *Zootaxa*, 2044: 46–60.

Moodley, Y. and Harley, E. H., 2005. Population structuring in mountain zebras (*Equus zebra*): the molecular consequences of divergent demographic histories. *Conservation Genetics*, 6: 953–968.

Moodley, Y. and Bruford, M. W., 2007. Molecular biogeography: towards an integrated framework for conserving pan-African biodiversity. *Plos One*, 2: e454.

Moore, A., Blenkinsop, T. and Cotterill, F., 2009. Southern African topography and erosion history: plumes or plate tectonics? *Terra Nova*, 21: 310–315.

Moore, A. E., Cotterill, F. and Eckardt, F., 2012. The evolution and ages of Makgadikgadi palaeo-lakes: consilient evidence from Kalahari drainage evolution south-central Africa. *South African Journal of Geology*, 115: 385–413.

Moritz, C., Patton, J. L., Schneider, C. J. and Smith, T. B., 2000. Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, 31: 533–563.

Morley, R. J., 2011. Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests, pp. 1-34 in *Tropical rainforest responses to climatic change*. Springer Publisher (M. Bush, J. Flenley and W. Gosling, Eds.), Berlin.

Muwanika, V. B., Nyakaana, S., Siegismund, H. R. and Arctander, P., 2003. Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, 91: 361–372.

Nersting, L. G. and Arctander, P., 2001. Phylogeography and conservation of impala and greater kudu. *Molecular Ecology*, 10: 711–719.

Nesi, N., Kadjo, B., Pourrut, X., Leroy, E., Shongo, C. P., Cruaud, C. and Hassanin, A., 2013. Molecular systematics and phylogeography of the tribe *Myonycterini* (Mammalia, Pteropodidae) inferred from mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution*, 66: 126–137.

Nicolas, V., Mboumba, J. F., Verheyen, E., Denys, C., Lecompte, E., Olayemi, A., Missoup, A. D., Katuala, P. and Colyn, M., 2008. Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography*, 35: 2074–2089.

Nicolas, V., Missoup, A. D., Denys, C., Kerbis Peterhans, J., Katuala, P., Couloux, A. and Colyn, M., 2011. The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *Journal of Biogeography*, 38: 191–207.

Ola, O. and Benjamin, E., 2019. Preserving biodiversity and ecosystem services in west african forest, watersheds, and wetlands: a review of incentives. *Forests*, 10: 479–497.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., et al., 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51: 933–938.

Osmaston, H. A. and Harrison, S. P., 2005. The Late Quaternary glaciation of Africa: a regional synthesis. *Quaternary International*, 138: 32–54.

Parr, C. L., Lehmann, C. E., Bond, W. J., Hoffmann, W. A. and Andersen, A. N., 2014. Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29: 205–213.

Pedersen, C. E. T., Albrechtsen, A., Etter, P. D., Johnson, E. A., Orlando, L., Chikhi, L., Siegismund, H. R and Heller, R., 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology & Evolution*, 2: 491–498.

Petružela, J., Šumbera, R., Aghová, T., Bryjová, A., Katakweba, A. S., Sabuni, C. A., Chitaukali, W. N. and Bryja, J., 2018. Spiny mice of the Zambezian bioregion-

phylogeny, biogeography and ecological differentiation within the *Acomys spinosissimus* complex. *Mammalian Biology*, 91: 79–90.

Pinton, A., Agnèse, J. F., Paugy, D. and Otero, O., 2013. A large-scale phylogeny of *Synodontis* (Mochokidae, Siluriformes) reveals the influence of geological events on continental diversity during the Cenozoic. *Molecular Phylogenetics and Evolution*, 66: 1027–1040.

Pitra, C., Hansen, A. J., Lieckfeldt, D. and Arctander, P., 2002. An exceptional case of historical outbreeding in African sable antelope populations. *Molecular Ecology*, 11: 1197–1208.

Plana, V., 2004. Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359: 1585–1594.

Pock Tsy, J. M. L., Lumaret, R., Mayne, D., Vall, A. O. M., Abutaba, Y. I., Sagna, M., Raoseta, S. O. R. and Danthu, P., 2009. Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* L.(Bombacoideae, Malvaceae). *Molecular Ecology*, 18: 1707–1715.

Pook, C. E., Joger, U., Stümpel, N. and Wüster, W., 2009. When continents collide: phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae). *Molecular Phylogenetics and Evolution*, 53: 792–807.

Portik, D. M., Bauer, A. M. and Jackman, T. R., 2011. Bridging the gap: western rock skinks (*Trachylepis sulcata*) have a short history in South Africa. *Molecular Ecology*, 20: 1744–1758.

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M-O, Blackburn, D. C. and Fujita, M. K., 2017. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26: 5245–5263.

Posada, D. and Buckley, T. R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53: 793–808.

QGIS Development Team, 2021. QGIS Geographic Information System. QGIS Association, available from http://www.qgis.org.

Rathbun, G. B., 2009. Why is there discordant diversity in sengi (Mammalia: Afrotheria: Macroscelidea) taxonomy and ecology? *African Journal of Ecology*, 47: 1–13.

Ratnam, J., Bond, W. J., Fensham, R. J., Hoffmann, W. A., Archibald, S., Lehmann, C. E., Anderson, M. T., Higgins, S. I. and Sankaran, M., 2011. When is a 'forest' a savanna, and why does it matter? *Global Ecology and Biogeography*, 20: 653–660.

Rebelo, L. M., McCartney, M. P. and Finlayson, C. M., 2010. Wetlands of Sub-Saharan Africa: distribution and contribution of agriculture to livelihoods. *Wetlands Ecology and Management*, 18: 557–572.

Rix, M. G., Edwards, D. L., Byrne, M., Harvey, M. S., Joseph, L. and Roberts, J. D., 2015. Biogeography and speciation of terrestrial fauna in the south-western Australian biodiversity hotspot. *Biological Reviews*, 90: 762–793.

Robert, C. and Chamley, H., 1987. Cenozoic evolution of continental humidity and paleoenvironment, deduced from the kaolinite content of oceanic sediments. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 60: 171–187.

Roberts, T. R., 1975. Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society*, 57: 249–319.

Roberts, E. M., Stevens, N. J., O'Connor, P. M., Dirks, P. H. G. M., Gottfried, M. D., Clyde, W. C., Armstrong, R. A., Kemp, A. I. S. and Hemming, S., 2012. Initiation of the western branch of the East African Rift coeval with the eastern branch. *Nature Geoscience*, 5: 289–294.

Robbins, C. B., 1978. The Dahomey Gap—a reevaluation of its significance as a faunal barrier to West African high forest mammals. *Bulletin of Carnegie Museum of Natural History*, 6: 168–74.

Rohland, N., Pollack, J. L., Nagel, D., Beauval, C., Airvaux, J., Pääbo, S. and Hofreiter, M., 2005. The population history of extant and extinct hyenas. *Molecular Biology and Evolution*, 22: 2435–2443.

Rokas, A. and Carroll, S., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*, 22: 1337–1344.

Russo, I. R. M., Chimimba, C. T. and Bloomer, P., 2006. Mitochondrial DNA differentiation between two species of *Aethomys* (Rodentia: Muridae) from southern Africa. *Journal of Mammalogy*, 87: 545–553.

Salzmann, U. and Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene*, 15: 190–199.

Sands, A. F., Matthee, S., Mfune, J. K. and Matthee, C. A., 2015. The influence of life history and climate driven diversification on the mtDNA phylogeographic structures of two southern African *Mastomys* species (Rodentia: Muridae: Murinae). *Biological Journal of the Linnean Society*, 114: 58–68.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., et al., 2005. Determinants of woody cover in African savannas. *Nature*, 438: 846–849.

Schmitt, T., 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4: 1–13.

Schreck III, C. J. and Semazzi, F. H., 2004. Variability of the recent climate of eastern Africa. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 24: 681–701.

Schuster, M., Roquin, C., Duringer, P., Brunet, M., Caugy, M., Fontugne, M., Mackaye, H. T., Vignaud, P. and Ghienne, J. F., 2005. Holocene lake Mega-Chad palaeoshorelines from space. *Quaternary Science Reviews*, 24: 1821–1827.

Scott, L., Steenkamp, M. and Beaumont, P. B., 1995. Palaeoenvironmental conditions in South Africa at the Pleistocene-Holocene transition. *Quaternary Science Reviews*, 14: 937–947.

Sepulchre, P., Ramstein, G., Fluteau, F., Schuster, M., Tiercelin, J. J. and Brunet, M., 2006. Tectonic uplift and Eastern Africa aridification. *Science*, 313: 1419–14

Sexton, G. J., Frere, C. H., Kalinganire, A., Uwamariya, A., Lowe, A. J., Godwin, I. D., Prentis, P. J. and Dieters, M. J., 2015. Influence of putative forest refugia and biogeographic barriers on the level and distribution of genetic variation in an African savannah tree, *Khaya senegalensis* (Desr.) A. Juss. *Tree Genetics & Genomes*, 11: 1–15.

Shafer, A. B., Cullingham, C. I., Cote, S. D. and Coltman, D. W., 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, 19: 4589–4621.

Shaw, P. A. and Thomas, D. S., 1996. The Quaternary palaeoenvironmental history of the Kalahari, southern Africa. *Journal of Arid Environments*, 32: 9–22.

Sheng, G. L., Soubrier, J., Liu, J. Y., Werdelin, L., Llamas, B., Thomson, V. A., et al., 2014. Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Molecular Ecology*, 23: 522–533.

Siesser, W. G. and Dingle, R. V., 1981. Tertiary sea-level movements around southern Africa. *The Journal of Geology*, 89: 83–96.

Smit, H. A., Robinson, T. J. and Jansen van Vuuren, B., 2007. Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). *Molecular Ecology*, 16: 2680–2692.

Smith, J. R., 2012. Spatial and temporal variation in the nature of Pleistocene pluvial phase environments across North Africa, pp. 35-47 in *Modern Origins*. Springer (J. J. Hublin and S. McPherron, Eds.), Dordrecht.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S. and Soltis, P. S., 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15: 4261–4293.

Soultan, A., Wikelski, M. and Safi, K., 2020. Classifying biogeographic realms of the endemic fauna in the Afro-Arabian region. *Ecology and Evolution*, 10: 8669–8680.

Stager, J. C. and Johnson, T. C., 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia*, 596: 5–16.

Stankiewicz, J. and de Wit, M., 2006. A proposed drainage evolution model for Central Africa – did the Congo flow east? *Journal of African Earth Sciences*, 44: 75–84.

Stokes, S., Thomas, D. S. G. and Washington, R., 1997. Multiple episodes of aridity in Africa south of the equator since the last interglacial cycle. *Nature*, 388: 154–158.

Šumbera, R., Krásová, J., Lavrenchenko, L. A., Mengistu, S., Bekele, A., Mikula, O. and Bryja, J., 2018. Ethiopian highlands as a cradle of the African fossorial root-rats (genus *Tachyoryctes*), the genetic evidence. *Molecular Phylogenetics and Evolution*, 126: 105–115.

Sungani, H., Ngatunga, B. P., Koblmüller, S., Mäkinen, T., Skelton, P. H. and Genner, M. J., 2017. Multiple colonisations of the Lake Malawi catchment by the genus *Opsaridium* (Teleostei: Cyprinidae). *Molecular Phylogenetics and Evolution*, 107: 256–265.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. and Cosson, J. F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7: 453–464.

Thieme, M. L., Abell, R., Burgess, N., Lehner, B., Dinerstein, E. and Olson, D., 2005. *Freshwater ecoregions of Africa and Madagascar: a conservation assessment.* Island Press, Washington.

Tolley, K. A., Braae, A. and Cunningham, M., 2010. Phylogeography of the Clicking Stream Frog *Strongylopus grayii* (Anura, Pyxicephalidae) reveals cryptic divergence across climatic zones in an abundant and widespread taxon. *African Journal of Herpetology*, 59: 17–32.

Trape, J. F., Chirio, L., Broadley, D. G. and Wüster, W., 2009. Phylogeography and systematic revision of the Egyptian cobra (Serpentes: Elapidae: *Naja haje*) species complex, with the description of a new species from West Africa. *Zootaxa*, 2236: 1–25.

Trauth, M. H., Deino, A., Center, B. G. and Strecker, M. R., 2005. Late Cenozoic moisture history of East Africa. *Science*, 309: 2051–2053.

Trauth, M. H., Larrasoaña, J. C. and Mudelsee, M., 2009. Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, 28: 399–411.

Ubangoh, R. U., Pacca, I. G. and Nyobe, J. B., 1998. Palaeomagnetism of the continental sector of the Cameroon Volcanic Line, West Africa. *Geophysical Journal International*, 135: 362–374.

Vamberger, M., Hofmeyr, M. D., Ihlow, F. and Fritz, U., 2018. In quest of contact: phylogeography of helmeted terrapins (*Pelomedusa galeata*, *P. subrufa* sensu stricto). *PeerJ*, 6: e4901.

van Daele, P. A., Verheyen, E., Brunain, M. and Adriaens, D., 2007. Cytochrome b sequence analysis reveals differential molecular evolution in African mole-rats of the chromosomally hyperdiverse genus *Fukomys* (Bathyergidae, Rodentia) from the Zambezian region. *Molecular Phylogenetics and Evolution*, 45: 142–157.

van der Walt, J., Nel, L. H. and Hoelzel, A. R., 2013. Differentiation at mitochondrial and nuclear loci between the blesbok (*Damaliscus pygargus phillipsi*) and bontebok (*D. p. pygargus*): implications for conservation strategy. *Conservation Genetics*, 14: 243–248.

Verheyen, E., Rüber, L., Snoeks, J. and Meyer, A., 1996. Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351: 797–805.

Visser, J. H., Bennett, N. C. and Jansen van Vuuren, B., 2018. Spatial genetic diversity in the Cape mole-rat, *Georychus capensis*: Extreme isolation of populations in a subterranean environment. *Plos One*, 13: e0194165.

Voelker, G., Bowie, R. C., Wilson, B. and Anderson, C., 2012. Phylogenetic relationships and speciation patterns in an African savanna dwelling bird genus (*Myrmecocichla*). *Biological Journal of the Linnean Society*, 106: 180–190.

Voelker, G., Marks, B. D., Kahindo, C., A'genonga, U., Bapeamoni, F., Duffie, L. E., Huntley, J. W., Mulotwa, E., Rosenbaum, S. A. and Light, J. E., 2013. River barriers and cryptic biodiversity in an evolutionary museum. *Ecology and Evolution*, 3: 536–545.

Voelker, G., Tobler, M., Prestridge, H. L., Duijm, E., Groenenberg, D., Hutchinson, M. R., Martin, A. D., Nieman, A., Roselaar, C. S. and Huntley, J. W., 2017. Three new species of *Stiphrornis* (Aves: Muscicapidae) from the Afro-tropics, with a molecular phylogenetic assessment of the genus. *Systematics and Biodiversity*, 15: 87–104.

Vrba, E., 1996. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate, pp. 385–424 in *Paleoclimate and Evolution*. Yale University Press (E. Vrba, G. Denton, T. Partidge and L. Burckle, Eds.), Connecticut.

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, A. S., Perkins, S. L. and Guralnick, R. P., 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *Plos One*, 2: e563.

Western, D., Groom, R. and Worden, J., 2009. The impact of subdivision and sedentarization of pastoral lands on wildlife in an African savanna ecosystem. *Biological Conservation*, 142: 2538–2546.

White, F., 1983. *The vegetation of Africa*. United Nations Educational, Scientific and Cultural Organization (Unesco, Ed.), Paris.

Willows-Munro, S. and Matthee, C. A., 2011. Linking lineage diversification to climate and habitat heterogeneity: phylogeography of the southern African shrew *Myosorex varius*. *Journal of Biogeography*, 38: 1976–1991.

Winter, S., Fennessy, J. and Janke, A., 2018. Limited introgression supports division of giraffe into four species. *Ecology and Evolution*, 8: 10156–10165.

Wondimu, T., Gizaw, A., Tusiime, F. M., Masao, C. A., Abdi, A. A., Gussarova, G., Popp, M., Nemomissa, S. and Brochmann, C., 2014. Crossing barriers in an extremely fragmented system: two case studies in the afro-alpine sky island flora. *Plant Systematics and Evolution*, 300: 415–430.

Xia, P., Radpour, R., Zachariah, R. and Fan, A., 2009. Simultaneous quantitative assessment of circulating cell-free mitochondrial and nuclear DNA by multiplex real-time PCR. *Genetics and Molecular Biology*, 32: 20–24.

Zimkus, B. M., Lawson, L. P., Barej, M. F., Barratt, C. D., Channing, A., Dash, K. M., et al., 2017. Leapfrogging into new territory: How Mascarene ridged frogs diversified across Africa and Madagascar to maintain their ecological niche. *Molecular Phylogenetics and Evolution*, 106: 254–269.

Zinner, D., Groeneveld, L. F., Keller, C. and Roos, C., 2009. Mitochondrial phylogeography of baboons (*Papio* spp.) – Indication for introgressive hybridization? *BMC Evolutionary Biology*, 9: 1–15.

2.6 Supporting Information

Text S2.1 Continental sub-Saharan Africa has a total of six classified areas with exceptional concentration of endemic species: 1) Guinean Forests of West Africa, comprising the upper Guinea and the Nigeria-Cameroon forest blocks; 2) Eastern Afromontane, including the Ethiopian Highlands, the Albertine Rift, and the Eastern Arc Mountains, as well as the Great Lakes; 3) the Coastal Forests of Eastern Africa, across Kenya and Tanzania; 4) the Succulent Karoo, a desert biome across the western coast of Namibia and South-Africa; 5) Cape Floristic Region, in South-Africa, holding about 10% of the world's flora biodiversity; and 6) the Maputaland-Pondoland-Albany, also in South-Africa, a floristic region situated bellow the Great Escarpment (Myers et al. 2000; Olson et al. 2001; Harrison and Noss 2017)[1]. From the literature search, a total of 99 research articles were based on taxa whose distributional range occupied a restricted area, namely across these defined hotspots of endemism (Table S2.1). These articles were not included in the analyses.

[1]Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. and Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature*, 403: 853-858.

Harrison, S. and Noss, R., 2017. Endemism hotspots are linked to stable climatic refugia. *Annals of Botany*, 119: 207-214.

Text S2.2 The QGIS plugin developed for the present work, named "linepattern", intersects a collection of lines with a user-defined grid size and derives a new general line, within each grid cell, depicting the major pattern found. The process starts by a simplification of the line within the square, considering only intersection points with the square and/or line end points within the square. Information about the number of lines in the square and the angle of each line considering the lower left line point as origin (Figure S2.2). Three scenarios are possible to produce the new simplified line: 1) the original lines only cross the grid cell at one side and a single line is generated based on the average position of the intersecting points with grid side and grid centroid; 2) two sides of the grid cell are intersected and average position on each side is used to create a single line; 3) three or four sides are intersected and three or four new lines are calculated based on average position of side intersection point and the centroid of the points within the grid square (Figure S2.3).

Table S2. 1 List of research articles on taxa with restricted distribution area, eliminate from our analysis (n = 99), per taxonomic group, and whether it corresponds to a described hotspot area of endemism.

Annotated restricted area	Endemic hostspot area	Number of articles						
		Mammals	Birds	Reptiles	Amphibians	Fish	Invertebrates	Plants
Afromontane region of Cameroon	—	–	2	–	2	–	–	–
Albertine Rift	Eastern afromontane	–	2	–	3	–	–	–
Cape Floristic Region	Cape Floristic Region	–	–	–	–	–	3	2
Coastal Forests of Eastern Africa	Coastal Forests of Eastern Africa	–	1	–	1	–	–	–
Drakensberg highlands (South-Africa)	—	–	–	–	1	–	–	1
Eastern afromontane	Eastern afromontane	8	8	3	1	–	–	3
Ethiopian highlands	Eastern afromontane	2	–	–	2	–	–	–
Guinean forest block	Guinean Forests of West Africa	–	–	–	–	–	1	–
Great Escarpment	—	–	–	1	–	–	–	–
Great Lakes	Eastern afromontane	–	–	–	–	7	2	–
Great Ruaha basin	—	–	–	–	–	1	–	–
Knersvlakte	—	–	–	–	–	–	–	1
Maputaland-Pondoland-Albany	Maputaland-Pondoland-Albany	1	–	–	–	–	–	–
Provinces of South-Africa (Cape and KwaZulu-Natal)	—	4	–	8	4	9	9	6

Table S2. 2 List of the 173 research articles used to perform the analysis of genetic discontinuities, ordered by taxonomic group (mammals, birds, reptiles, amphibians, fish, invertebrates, and plants), then by year of publication, and author's name, in alphabetical order.

Year of publication	Author	Doi	Taxonomic Group	Target taxa
1996	Matthee and Robinson	10.1038/hdy.1996.74	Mammals	*Pronolagus rupestris*
1997	Matthee and Robinson	10.1023/A:1027331727034	Mammals	*Pedetes capensis; P. surdaster*
1999	Arctander et. al.	10.1093/oxfordjournals.molbev.a026085	Mammals	*Alcelaphus buselaphus; Connochaetes taurinus; Damaliscus lunatus*
2001	Flagstad et. al.	10.1098/rspb.2000.1416	Mammals	*Alcelaphus buselaphus*
2001	Nersting and Arctander	10.1046/j.1365-294x.2001.01205.x	Mammals	*Tragelaphus strepsiceros*
2002	Comstock et. al.	10.1046/j.1365-294X.2002.01615.x	Mammals	*Loxodonta africana*
2002	Van Hooft et. al.	10.1046/j.1365-294X.2002.01429.x	Mammals	*Syncerus caffer*
2003	Muwanika et. al.	10.1038/sj.hdy.6800341	Mammals	*Phacochoerus africanus*
2004	Alpers et. al.	10.1111/j.1365-294X.2004.02204.x	Mammals	*Hippotragus equinus*
2005	Herron et. al.	10.1111/j.1365-294X.2005.02630.x	Mammals	*Xerus inauris; X. princeps*
2005	Moodley and Harley	10.1007/s10592-005-9083-8	Mammals	*Equus zebra*
2005	Rohland et. al.	10.1093/molbev/msi244	Mammals	*Crocuta crocuta*
2006	Lorenzen et. al.	10.1093/jhered/esj012	Mammals	*Aepyceros melampus*
2006	Lorenzen et. al.	10.1111/j.1365-294X.2006.03059.x	Mammals	*Kobus ellipsiprymnus*
2006	Masembe et. al.	10.1007/s10592-005-9066-9	Mammals	*Oryx beisa*
2006	Pitra et al.	10.1007/s10344-005-0026-y	Mammals	*Hippotragus niger*
2006	Russo et. al.	10.1644/05-MAMM-A-222R3.1	Mammals	*Aethomys ineptus*
2007	Moodley et al.	10.1371/journal.pone.0000454	Mammals	*Tragelaphus scriptus*
2007	van Daele et. al.	10.1016/j.ympev.2007.04.008	Mammals	*Fukomys damarensis; F. micklemi*
2008	Dehghani et. al.	10.1111/j.1469-7998.2008.00502.x	Mammals	*Ichneumia albicauda*
2008	Lamb et. al.	10.3161/150811008X331063	Mammals	*Otomops martiensseni*
2008	Lorenzen et. al.	10.1007/s10592-007-9375-2	Mammals	*Nanger granti*

2008	Moulini et. al.	10.1111/j.1365-294X.2007.03610.x	Mammals	*Mastomys huberti*
2008	Nicolas et al.	10.1111/j.1365-294X.2008.03974.x	Mammals	*Praomys rostratus; P. tullbergi*
2008	Nicolas et. al.	10.1111/j.1365-2699.2008.01950.x	Mammals	*Lemniscomys striatus*
2009	Brouat et. al.	10.1111/j.1365-2699.2009.02184.x	Mammals	*Mastomys erythroleucus*
2009	Zinner et. al.	10.1186/1471-2148-9-83	Mammals	*Papio* spp.
2010	Bryja et. al.	10.1111/j.1365-294X.2010.04847.x	Mammals	*Praomys daltoni* spp. complex
2010	Colyn et. al.	10.11646/zootaxa.2637.1.1	Mammals	*Philantomba maxwelli; P. walteri*
2010	Lorenzen et. al.	10.1111/j.1365-2699.2009.02207.x	Mammals	*Tragelaphus oryx*
2010	Nicolas et al.	10.1111/j.1096-3642.2009.00602.x	Mammals	*Praomys misonnei; P. tullbergi; P. misonnei*
2010	Russo et. al.	10.1186/1471-2148-10-307	Mammals	*Micaelamys namaquensis*
2010	Smit et. al.	10.1111/j.1365-2699.2009.02249.x	Mammals	*Elephantulus rupestris*
2010	van Vuuren et. al.	10.3957/056.040.0114	Mammals	*Hippotragus niger*
2011	Charruau et. al.	10.1111/j.1365-294X.2010.04986.x	Mammals	*Acinonyx jubatus*
2011	Edwards et. al.	10.1111/j.1095-8312.2010.01583.x	Mammals	*Otomys unisulcatus*
2011	Engelbrecht et. al.	10.1111/j.1095-8312.2011.01696.x	Mammals	*Otomys irroratus*
2011	Ishida et. al.	10.1371/journal.pone.0020642	Mammals	*Loxodonta africana*
2011	Kennis et. al.	10.1111/j.1096-3642.2011.00733.x	Mammals	*Praomys* spp.
2011	Nicolas et al.	10.1111/j.1365-2699.2010.02399.x	Mammals	*Praomys misonnei*
2011	Willows-Munro and Matthee	10.1111/j.1365-2699.2011.02543.x	Mammals	*Myosorex varius*
2012	Castiglia et. al.	10.1111/j.1439-0469.2011.00627.x	Mammals	*Rhabdomys dilectus*
2012	Du Toit et. al.	10.1016/j.ympev.2012.05.036	Mammals	*Rhabdomys pumilio*
2012	Dumbacher et. al.	10.1371/journal.pone.0032410	Mammals	*Macroscelides proboscideus*
2012	Marsden et. al.	10.1111/j.1365-294X.2012.05477.x	Mammals	*Lycaon pictus*
2012	Osmers et. al.	10.1016/j.mambio.2011.08.003	Mammals	*Oryx gazella*
2013	Colangelo et. al.	10.1111/bij.12013	Mammals	*Mastomys natalensis*
2013	Dobigny et. al.	10.1371/journal.pone.0077815	Mammals	*Arvicanthis niloticus*

2013	Haus et. al.	10.1002/ajp.22113	Mammals	*Chlorocebus* spp.
2013	Jacobs et. al.	10.1371/journal.pone.0082614	Mammals	*Rhinolophus damarensis; R. darlingi*
2013	Nesi et al.	10.1016/j.ympev.2012.09.028	Mammals	*Megaloglossus azagny; Megaloglossus woermanni; Myonycteris torquata; Myonycteris leptodon*
2013	Smitz et. al.	10.1371/journal.pone.0056235	Mammals	*Syncerus caffer*
2013	van der Walt et. al.	10.1007/s10592-012-0435-x	Mammals	*Damaliscus pygargus*
2014	Bryja et. al.	10.1111/jbi.12195	Mammals	*Praomys delectorum*
2014	Bryja et. al.	10.1186/s12862-014-0256-2	Mammals	*Mus minutoides; M. setulosus*
2014	Jacquet et. al.	10.1111/zsc.12039	Mammals	*Crocidura eburnea; C. obscurior*
2014	Sheng et. al.	10.1111/mec.12576	Mammals	*Crocuta crocuta*
2014	Sonet et. al.	10.1111/jzs.12066	Mammals	*Crossarchus alexandri; C. platycephalus*
2015	Bohoussou et. al.	10.1111/jbi.12570	Mammals	*Malacomys edwardsi; M. longipes*
2015	Hassanin et. al.	10.1016/j.crvi.2014.12.003	Mammals	*Scotonycteris zenkeri*
2015	Jacquet et. al.	10.1186/s12862-015-0344-y	Mammals	*Crocidura* spp.
2015	le Grange et. al.	10.1080/15627020.2015.1021178	Mammals	*Rhabdomys dilectus; R. pumilio*
2015	McDonough et. al.	10.1111/mec.13374	Mammals	*Gerbilliscus leucogaster*
2015	Sands et. al.	10.1111/bij.12397	Mammals	*Mastomys coucha; M. natalensis*
2015	Stoffel et. al.	10.1111/mec.13179	Mammals	*Hippopotamus amphibius*
2016	Bertola et al.	10.1038/srep30807	Mammals	*Panthera leo*
2016	Gaubert et. al.	10.1111/mec.13886	Mammals	*Manis tricuspis*
2016	Hassanin et. al.	10.1016/j.crvi.2016.09.005	Mammals	*Eilodon helvum; Epomops franqueti; E. buettikoferi; Hypsignathus monstrosus; Micropteropus pusillus; Nanonycteris veldkampii; Rousettus aegyptiacus*
2016	Lobon et. al.	10.1093/gbe/evw124	Mammals	*Pan troglodytes*
2016	Mikula et al.	10.1111/zsc.12179	Mammals	*Saccostomus campestris; S. mearnsi*
2017	Aghová et. al.	10.1111/jbi.13017	Mammals	*Gerbilliscus* spp.
2017	Carlen et. al.	10.1016/j.ympev.2017.05.012	Mammals	*Rhynchocyon* spp.
2017	Mazoch et. al.	10.1515/mammalia-2017-0001	Mammals	*Aethomys chrysophilus*

2018	Anco et. al.	10.1080/24701394.2017.1307973	Mammals	*Panthera pardus*
2018	Bryja et al.	10.1016/j.mambio.2018.08.006	Mammals	*Heliophobius argenteocinereus*
2018	Patterson et. al.	10.7717/peerj.4864	Mammals	*Otomops harrisoni*
2018	Pedersen et. al.	10.1038/s41559-017-0453-7	Mammals	*Equus quagga*
2018	Petruzela et al.	10.1016/j.mambio.2018.03.012	Mammals	*Acomys* spp.
2018	Šumbera et. al.	10.1016/j.ympev.2018.04.003	Mammals	*Tachyoryctes splendens*
2018	Visser et. al.	10.1371/journal.pone.0194165	Mammals	*Georychus capensis*
2018	Winter et. al.	10.1002/ece3.4490	Mammals	*Giraffa camelopardalis*
1999	Beresford and Cracraft		Birds	*Stiphrornis* spp.
2004	Bowie et. al.	10.1016/j.ympev.2004.04.013	Birds	*Nectarinia olivacea; N. obscura*
2006	Bowie et. al.	10.1016/j.ympev.2005.06.001	Birds	*Pogonocichla stellata*
2010	Marks	10.1016/j.ympev.2009.10.027	Birds	*Hylia prasina*
2011	Fuchs et. al.	10.1111/j.1365-2699.2011.02545.x	Birds	*Lanius collaris* spp. complex
2011	Fuchs et. al.	10.1186/1471-2148-11-117	Birds	*Phyllastrephus debilis*
2011	Miller et. al.	10.1007/s10592-010-0149-x	Birds	*Struthio camelus*
2012	Voelker et. al.	10.1111/j.1095-8312.2012.01856.x	Birds	*Myrmecocichla* spp.
2014	Engel et. al.	10.1111/ibi.12145	Birds	*Cercococcyx montanus*
2015	Fuchs and Bowie	10.1016/j.ympev.2015.03.011	Birds	*Campethera caroli; C. nivosa*
2016	Huntley and Voelker	10.1016/j.ympev.2016.04.002	Birds	*Criniger barbatus; C. calurus; C. olivaceus*
2017	Voelker et. al.	10.1080/14772000.2016.1226978	Birds	*Stiphrornis erythrothorax*
2018	Fuchs et. al.	10.1111/zsc.12274	Birds	*Dicrurus ludwigii sensu lato*
2018	Fuchs et. al.	10.11646/zootaxa.4438.1.4	Birds	*Dicrurus atripennis; D. ludwigii*
2018	Haworth et. al.	10.7717/peerj.5000	Birds	*Promerops cafer; P. gurneyi*
2018	Huntley et. al.	10.1093/zoolinnean/zlx086	Birds	*Criniger barbatus*
2002	Matthee and Flemming	10.1046/j.0962-1083.2001.01458.x	Reptiles	*Agama atra*
2007	Daniels et. al.	10.1016/j.ympev.2007.08.010	Reptiles	*Chersina angulata*

2009	Eaton et. al.	10.1016/j.ympev.2008.11.009	Reptiles	*Osteolaemus tetraspis*
2009	Pook et. al.	10.1016/j.ympev.2009.08.002	Reptiles	*Echis ocellatus*
2009	Trape et. al.	10.5281/zenodo.190424	Reptiles	*Naja haje* complex
2010	Fritz et. al.	10.1111/j.1439-0469.2010.00565.x	Reptiles	*Stigmochelys pardalis*
2010	Leaché and Fujita	10.1098/rspb.2010.0662	Reptiles	*Hemidactylus fasciatus*
2010	Vargas-Ramírez et. al.	10.1016/j.ympev.2010.03.019	Reptiles	*Pelomedusa subrufa*
2010	Wong et. al.	10.3854/crm.5.007.subrufa.v1.2008	Reptiles	*Pelomedusa subrufa*
2011	Portik et. al.	10.1111/j.1365-294X.2011.05047.x	Reptiles	*Trachylepis sulcata*
2013	Barlow et. al.	10.1111/mec.12157	Reptiles	*Bitis arietans*
2014	Diedericks and Daniels	10.1016/j.ympev.2013.10.015	Reptiles	*Cordylus cordylus*
2015	Lamb and Bauer	10.1080/15627020.2000.11407192	Reptiles	*Pachydactylus rugosus*
2016	Dowell et. al.	10.1016/j.ympev.2015.10.004	Reptiles	*Varanus niloticus*
2017	Busschau et. al.	10.1016/j.zool.2016.11.005	Reptiles	*Acontias breviceps; A. gracilicauda*
2018	Freitas et. al.	10.1016/j.ympev.2018.04.020	Reptiles	*Mochlus sundevalli-afer* spp. complex
2018	Vamberger et. al.	10.7717/peerj.4901	Reptiles	*Pelomedusa galeata*
2018	Weinell and Bauer	10.1016/j.ympev.2017.11.014	Reptiles	*Trachylepis varia*
2018	Wüster et. al.	10.11646/zootaxa.4455.1.3	Reptiles	*Naja melanoleuca* complex
2004	Evans et. al.	10.1016/j.ympev.2004.04.018	Amphibians	*Xenopus tropicalis/ epitropicalis*
2010	Tolley et. al.	10.1080/04416651003744943	Amphibians	*Strongylopus grayii*
2013	Lawson	10.1111/mec.12229	Amphibians	*Hyperolius substriatus*
2015	Furman et. al.	10.1111/mec.13076	Amphibians	*Xenopus laevis sensu lacto*
2017	Bittencourt-Silva et. al.	10.1016/j.ympev.2017.06.022	Amphibians	*Hyperolius substriatus*
2017	Manthey et. al.	10.1093/biolinnean/blx106	Amphibians	*Amietia nutti*
2017	Portik et. al.	10.1111/mec.14266	Amphibians	*Scotobleps gabonicus*
2017	Zimkus et al.	10.1016/j.ympev.2016.09.018	Amphibians	*Ptychadena mascareniensis*
2018	Barratt et. al.	10.1111/mec.14862	Amphibians	*Afrixalus delicatus; Afrixalus fornasini; Arthroleptis xenodactyloides; Leptopelis flavomaculatus*

2018	Jongsma et al.	10.1016/j.ympev.2017.12.006	Amphibians	*Amnirana albolabris; A. Galamensis*
2007	Kramer et. al.	10.1080/00222930701250987	Fish	*Marcusenius macrolepidotus* spp. complex
2011	Bezault et. al.	10.1186/1471-2156-12-102	Fish	*Oreochromis niloticus*
2011	Goodier et. al.	10.1371/journal.pone.0028775	Fish	*Hydrocynus* spp.
2013	Day et. al.	10.1093/sysbio/syt001	Fish	*Synodontis* spp.
2013	Pinton et. al.	10.1016/j.ympev.2012.12.009	Fish	*Synodontis* spp.
2015	Egger et. al.	10.1007/s10750-014-1919-0	Fish	*Pseudocrenilabrus philander*
2016	Kinde et al.	10.1111/jfb.12899	Fish	*Coptodon guineensis* spp. complex*; C. zillii; Sarotherodon melanotheron; S. nigrippinis*
2016	Riddin et. al.	10.3897/zookeys.641.10434	Fish	*Mesobola* spp.
2017	Sungani et. al.	10.1016/j.ympev.2016.09.027	Fish	*Opsaridium* spp.
2006	Sezonlin et. al.	10.1080/00379271.2006.10697466	Invertebrates	*Busseola fusca*
2003	Daniels	10.1651/C-2383	Invertebrates	*Potamonautes perlatus*
2003	Dejong et. al.	10.1046/j.1365-294X.2003.01977.x	Invertebrates	*Biomphalaria pfeifferi*
2006	Daniels et. al.	10.1111/j.1365-2699.2006.01537.x	Invertebrates	*Potamonautes perlatus*
2009	Daniels et. al.	10.1111/j.1095-8312.2009.01205.x	Invertebrates	*Peripatopsis moseleyi*
2010	Daniels et. al.	10.1111/j.1469-7998.2010.00722.x	Invertebrates	*Peripatopsis moseleyi*
2014	Phiri and Daniels	10.1111/zoj.12103	Invertebrates	*Potamonautes perlatus*
2016	Daniels et. al.	10.1111/cla.12154	Invertebrates	*Opisthopatus cinctipes*
2006	Kadu et. al.	10.1016/j.sajb.2005.12.007	Plants	*Sclerocarya birrea*
2007	Assefa et. al.	10.1038/sj.hdy.6800974	Plants	*Arabis alpina*
2007	Enrich et. al.	10.1111/j.1365-294X.2007.03299.x	Plants	*Arabis alpina*
2007	Kebede et. al.	10.1111/j.1365-294X.2007.03232.x	Plants	*Lobelia giberroa*
2009	Gomez et. al.	10.1186/1471-2148-9-167	Plants	*Coffea canephora*
2009	Pock Tsy et. al.	10.1111/j.1365-294X.2009.04144.x	Plants	*Adansonia digitata*
2010	Anthony et. al.	10.1007/s00606-009-0255-8	Plants	*Coffea* spp.
2011	Allal et. al.	10.1038/hdy.2011.5	Plants	*Vitellaria paradoxa*

2011	Desamore et. al.	10.1111/j.1365-2699.2010.02387.x	Plants	*Erica arborea*
2011	Koffi et. al.	10.3732/ajb.1000220	Plants	*Santiria trimera*
2011	Liu et. al.	10.1093/aob/mcr231	Plants	*Eleusine* spp.
2012	Odee et. al.	10.1038/hdy.2012.52	Plants	*Acacia senegal*
2013	Budde et. al.	10.1038/hdy.2013.21	Plants	*Symphonia globulifera*
2013	Duminil et. al.	10.1186/1471-2148-13-195	Plants	*Erythrophleum ivorense; E. suaveolens*
2013	Gisaw et. al.	10.1016/j.flora.2013.07.007	Plants	*Erica arborea; E. trimera*
2013	Kadu et. al.	10.1093/aob/mcs235	Plants	*Prunus africana*
2014	Daïnou et. al.	10.1038/hdy.2014.5	Plants	*Milicia excelsa; M. regia*
2014	Linder et. al.	10.1600/036364414X680906	Plants	*Tenaxia disticha*
2014	Wondimu et. al.	10.1007/s00606-013-0892-9	Plants	*Carduus schimperi; Trifolium cryptopodium*
2015	Castro et. al.	10.1093/aob/mcv001	Plants	*Cyperus esculentus*
2015	Chen et. al.	10.1007/s11434-015-0832-x	Plants	*Haplocarpha rueppelii*
2015	Sexton et. al.	10.1007/s11295-015-0933-3	Plants	*Khaya senegalensis*
2016	Chartier et. al.	10.12705/656.6	Plants	*Delphinium* spp.
2016	Demenou et la.	10.1111/jbi.12688	Plants	*Distemonanthus benthamianus*
2016	Gizaw et al.	10.1007/s00035-015-0162-2	Plants	*Carex monostachya*
2017	Gichira et. al.	10.1007/s11295-017-1156-6	Plants	*Hagenia abyssinica*
2017	Mairal et. al.	10.1038/srep45749	Plants	*Canarina eminii*
2018	Demenou et. al.	10.1038/s41437-017-0035-0	Plants	*Terminalia superba*
2018	Lompo et. al.	10.1093/jhered/esy047	Plants	*Parkia biglobosa*
2018	Stångberg et. al.	10.1016/j.sajb.2018.07.018	Plants	*Hirpicium* spp.

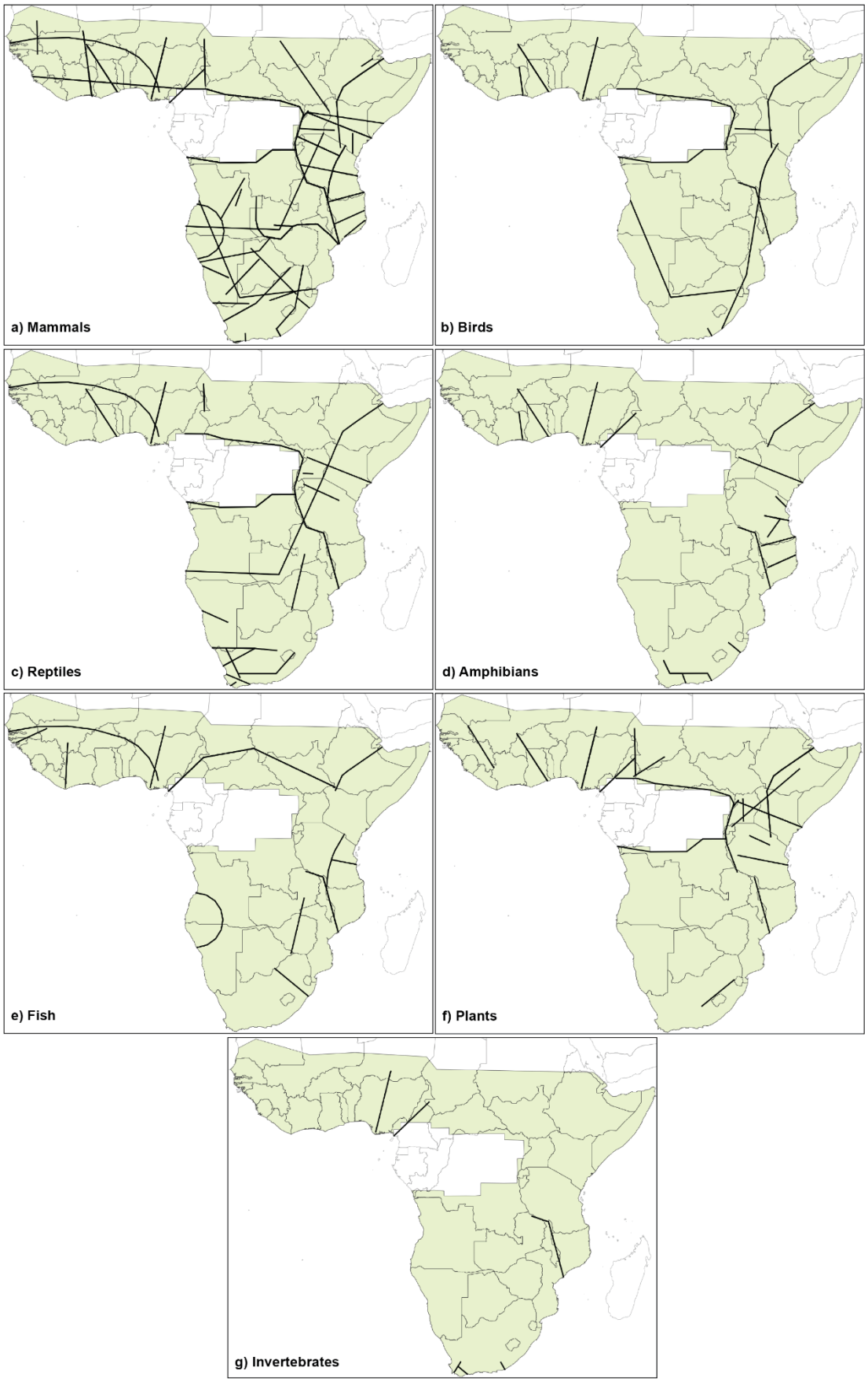


Figure S2. 1 Maps showing the annotated genetic discontinuities (before applying the plug-in) across sub-Saharan Africa, per taxonomic group: **a)** Mammals; **b)** Birds; **c)** Reptiles; **d)** Amphibians; **e)** Fish; **f)** Plants; **g)** Invertebrates.

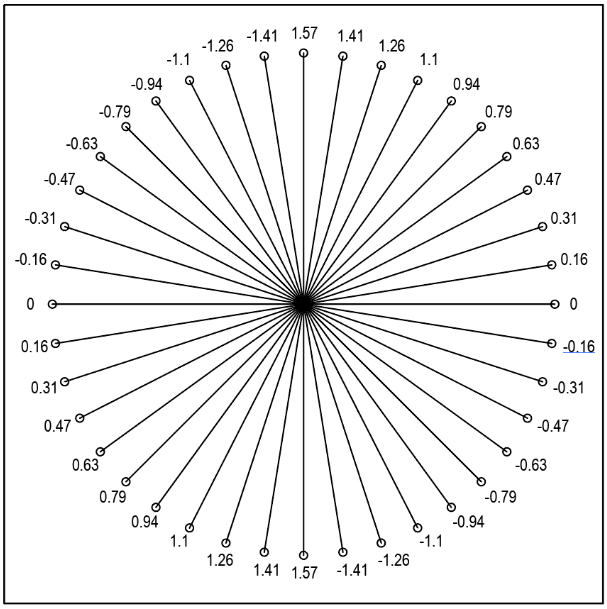


Figure S2. 2 Calculation of line angle (radians) within each grid square, by the plug-in. Orientation of the line is always considered from left to right and angles are calculated in relation to the east-west line.

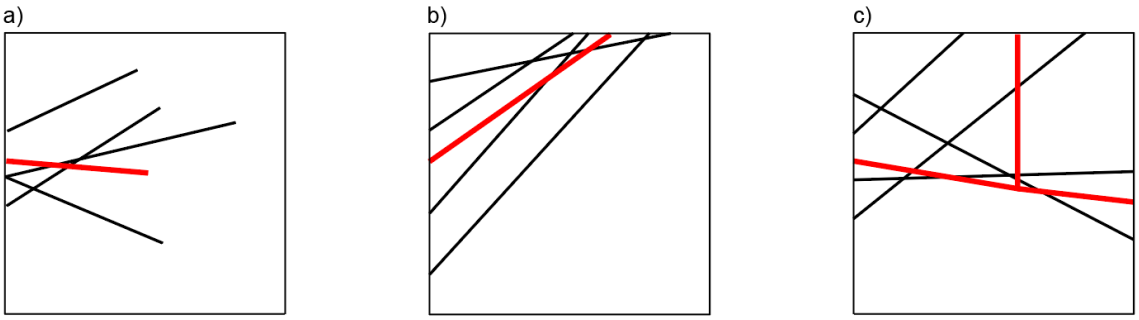


Figure S2. 3 Illustration on how the plug-in processes the generation a new simplified line from a set of intersecting lines. Three scenarios are possible: **a)** one side of the grid square is intersected; **b)** two sides are intersected; and **c)** three or four sides are intersected. Average position on each side and centroids are used to produce the new lines, depending on the case. Black lines represent the original lines intersected with the grid square and the red lines represent the new generated, simplified line.

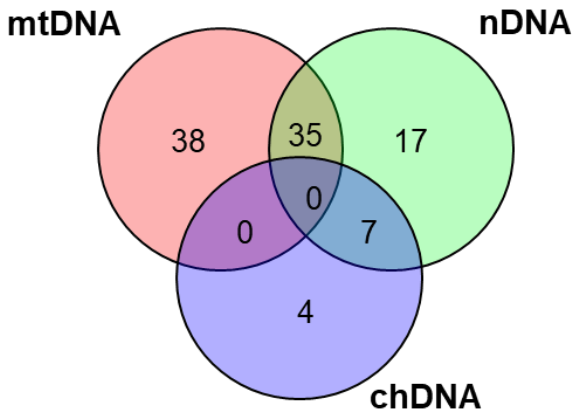


Figure S2. 4 Venn diagram showing the type of molecular marker applied on the 173 research articles used to analyse the genetic discontinuities. mtDNA – mitochondrial DNA; nDNA – nuclear DNA (combining sequences and microsatellite loci); and chDNA – chloroplast DNA.

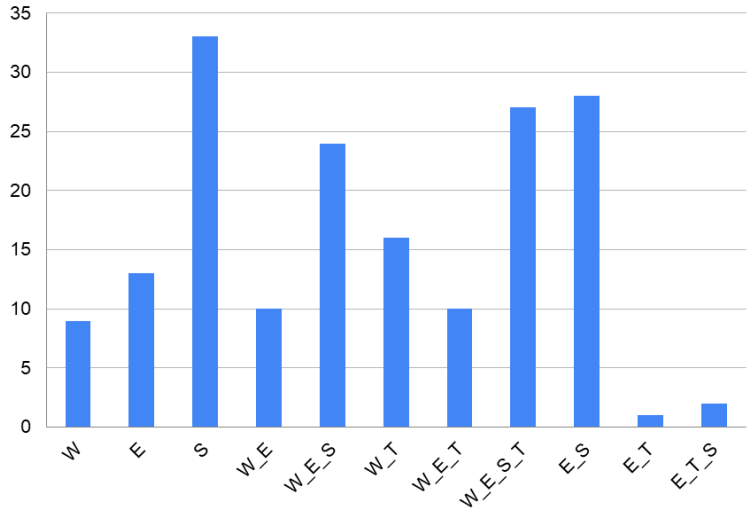


Figure S2. 5 Graph showing the number of research articles sampling across all possible combinations of the main regions considered across sub-Saharan Africa: W – West; E – East; S – South (as defined by Hewitt 2004); and T – Tropical Forest belt, delimited by the African Congolian Forest biogeographic region (as defined by Linder et al. 2012).

CHAPTER 3

Population analyses of the roan antelope

Study I

Evolutionary history of the roan antelope across its African range

Margarida Gonçalves[1,2], Hans R. Siegismund[3], Bettine Jansen van Vuuren[4], Nuno Ferrand[1,2,4], and Raquel Godinho[1,2,4]

Authors' affiliations: [1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, Portugal; [2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto; [3]Department of Biology, University of Copenhagen, Denmark; [4]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South-Africa

3.1 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 and 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 (Vrba 1996; Dupont 2011). Specifically, a notable contributing factor that drove much of the speciation events is habitat fluctuations during Pleistocene climatic cycles (DeMenocal 2004; Lorenzen, Heller and Siegismund 2012), combined with continental-wide geomorphological changes throughout the Pliocene and Pleistocene (Goudie 2005; Trauth, Larrasoaña and Mudelsee 2009; Badgley 2010).

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 refuge populations during periods of isolation, as well as the accumulation of genetic diversity because of species' persistence. During interpluvials, recolonization of newly available habitats allowed for the spread of new mutations and 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 been documenting congruent spatial patterns among African ungulates (see e.g. Arctander, Johansen and Coutellec-Vreto 1999; Lorenzen, et al. 2010; Smitz et al. 2013), largely related to Pleistocene climatic oscillations. Based on their own work and data available for other species, Lorenzen and co-workers (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) (Trauth et al. 2009; Moorley and Kingdon 2013). However, for species with pan-African ranges, spatial genetic patterns remain poorly documented and understood, largely because of the paucity of studies that include data covering the species' entire distributions.

Roan (*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 and 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 roan antelope (Matthee and Robinson 1999; Alpers et al. 2004). Also, the spatial genetic patterns that emerged from the partial coverage of this antelope's range challenged the validity of the recognized subspecies (Ansell 1972), and instead proposed the presence of two Evolutionary Significant Units (ESU). The first ESU is situated in West Africa and included the subspecies *H. e. koba* in the north-west. The second ESU included *H. e. langheldi* in East Africa, *H. e. cottoni* in south-central Africa, and *H. e. equinus* in the southern Africa. No or very limited samples were available for *H. e. charicus* in north-central and *H. e. bakeri* in north-east Africa. Although important, studies that suffer from incomplete geographic sampling may lead to only a partial or even erroneous understanding of spatial genetic patterns and hence, also the spatial drivers of these patterns.

Other inherent characteristics may influence the evolutionary history and demographic trends present in roan antelopes. Although having a continental range, they 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 fifteen females and calves (Chardonnet

and Crosmary 2013). Of some concern is that 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 clear that these animals are more common throughout West and Central Africa, whereas across some parts of its eastern and southern distribution, some populations are considered rare and even virtually extinct (East and IUCN SSC Antelope Specialist Group 1999; Chardonnet and Crosmary 2013).

In this regard, the roan antelope is an ideal system for a more comprehensive review of its population genetic parameters and accurate number of distinct spatial units, with direct implications towards management plans. Based on previous work (Matthee and Robinson 1999; Alpers et al. 2004; Lorenzen et al. 2012), we hypothesize that this species harbours deep genetic signatures of Pleistocene climatic oscillations, with populations having uneven genetic diversity and concerning demographic trends. To test this hypothesis, we provide the most comprehensive assessment of this species' phylogeography to date, based on samples from across the entire pan-African range of the roan antelope. We generated data from a large suite of microsatellite markers and complete mitogenomes, and include both contemporary and historical samples. Our results give a genetic context for the recognized subspecies and highlight contact zones, providing some suggestions for this iconic species' management and conservation. We address the evolutionary history against the backdrop of the climatic and geomorphological history of the African continent throughout the Pleistocene, providing estimates on the timing of vicariant events, and importance of geographical barriers and their permeability.

3.2 Materials and Methods

3.2.1 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 S3.1). For the contemporary samples, total genomic DNA was extracted from muscle samples using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germany) 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 S3.1 for details). For both methods, blank controls were routinely included to test for contamination. Eluted DNA concentrations were quantified by fluorometry using the Qubit dsDNA HS - High Sensitivity - Assay Kit (Thermo Fisher Scientific, USA). Final

DNA concentrations averaged 40 ng/µl for contemporary samples (range between 1 and 120 ng/µl), and 25 ng/µl for historical samples (range between 0.3 and 100 ng/µl).

3.2.2 Microsatellite genotyping

Vaz Pinto et al. (2015) developed a suite of 57 microsatellite markers for the sable antelope (*Hippotragus niger*), the sister taxon to the roan antelope. We followed their methodology and optimized 54 of these markers for use in roan antelope samples (see Text S3.2 and Table S3.2 for details). Blank controls were routinely included while duplicate (for contemporary) and triplicate (for historic) samples were amplified to minimize possible amplification errors. Using GeneMapper v.4.0 (Applied Biosystems, USA), 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 less than 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 v.6.5.3 (Peakall and 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.

3.2.3 Nuclear data analyses

Spatial genetic structure (SGS) was first assessed for the whole dataset using Structure v.2.3.4 (Pritchard, Stephens and Donnelly 2000; Falush, Stephens and Pritchard 2003; Hubisz et al. 2009); 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 MCMC iterations following a burn-in period of 0.5 million steps. Structure Harvester Web v.0.6.94 (Earl and VonHoldt 2012) was used to plot the likelihood for each K using the mean likelihood and delta-K statistics (Evanno, Regnaut and Goudet 2005). Results from this first analysis (run 1) revealed the highest delta-K value for K = 2 (Figure S3.1b). However, this statistic might not be reliable under complex evolutionary scenarios (Janes et al. 2017), and additional genetic substructure for subsequent values of K (Figure S3.1c), prompted us to partition the data according to

their assignment probability (q) for K = 2, as suggested by Pritchard, Wen and Falush, (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 3.1a). 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. 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 v.3.5 (Excoffier and 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 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 v.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 G'_{ST} statistic (Hedrick 2005) (G'_{ST} corrects F_{ST} biases for microsatellite loci). We also performed an analysis of molecular variance (AMOVA) in Arlequin with 1,000 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 and Chevalet 2014) implemented in R v.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: $1x10^{-3}$ and $1x10^{-4}$ mutations/locus/year (Ritz et al. 2000; Qi et al. 2015) 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 Markov Chain Monte Carlo (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 and Crosmary 2013).

3.2.4 Whole mitochondrial sequencing and assembly

For contemporary samples, double-stranded DNA libraries were prepared using Nextera XT DNA Library Prep Kit (Illumina, USA), following the modifications described in Tan and Mikheyev (2016). After a sonication step, size-selection of DNA fragments of ca. 350 bp was performed using AMPure XP bead clean-up (Beckman Coulter, USA). For historical samples, DNA library preparation followed Meyer and Kircher (2010) with modifications proposed by Kircher and co-workers (2012). Amplification and purification of these DNA libraries followed Dabney and Meyer (2012). We performed a target sequencing capture, using four fragments of ca. 5,000 bp as bait, covering the entire mitochondrial genome of 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 roan antelope (see Text S3.3 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, USA) on a single cell using the 250 bp paired-end sequencing protocol.

Raw sequences were demultiplexed after adapter trimming using Trimmomatic v.0.36 (Bolger, Lohse and Usadel 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 *mem* from the mapper BWA v.0.7.16 (Li 2013). Base-calling was performed using SAMTools v.1.5 and BCFTools v.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 Abyss-pe v.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)).

3.2.5 Mitochondrial data analysis

The final dataset was aligned using Muscle v.3.8.31 (Edgar 2004). Coding regions were translated using DnaSP v.5.0 (Librado and Rozas 2009); no non-sense mutations were detected. The alignment for the hyper-variable 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, Forster and Röhl 1999) was constructed using PopART v.1.7 (http://popart.otago.ac.nz). Arlequin was used to assess mitochondrial genetic diversity within each lineage and nuclear group through both haplotype (H_d) and nucleotide (π) diversities, as well as mean pairwise distances (MPD), 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 1,000 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 African oryx or beisa oryx (*Oryx beisa*; NC_020793.1), the scimitar-horned oryx (*Oryx dammah*; NC_016421.1), the gemsbock (*Oryx gazella*; NC_016422.1), the Arabian oryx (*Oryx leucoryx*; NC_020732.1), and the addax (*Addax nasomaculatus*; NC_020674.1). The remaining three outgroups belonged to species from the Alcelaphinae, namely 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 of roan's mitochondrial lineages (see Table S3.1 for details). PartitionFinder v.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; 1st and 2nd codon-position of the protein-coding genes; 3rd codon-position of the protein-coding genes. Bayesian phylogenetic analyses were performed in

Beast v.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. Calibration ages resulting from Bibi's (2013) mitochondrial phylogeny for Bovidae were used as priors for the root node and main nodes of the Hippotraginae and Alcelaphinae sub-families (Table S3.3). 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 v.1.7.1 (Rambaut et al. 2018) considering only models whose statistics had an effective sample size (ESS) value above 1,000. Maximum clade credibility (MCC) trees were estimated with TreeAnnotator v.2.5.1, discarding a 10% burn-in. The consensus tree was visualized in FigTree v.1.4.4.

To test for signatures of demographic expansion within each roan lineage, we used the whole dataset and estimated neutrality tests based on Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) with 1,000 permutations, implemented in Arlequin.

3.3 Results

3.3.1 Nuclear data

The final dataset included 131 individuals from 70 localities across sub-Saharan Africa, genotyped for 43 autosomal microsatellites (Table S3.1 and Table S3.2). Ten of the initial loci were excluded from our analyses due to 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 S3.4). No significant deviations from LD were observed for any pair of loci across groups.

3.3.1.1 Population structure

Our initial clustering analysis (run 1), retrieved K = 2 as the most likely number of groups, based on delta-K statistic (Figure S3.1b). The first group, with assignment probability > 0.5 to the first cluster, corresponded to animals in West Africa, from the Central African Republic (CAR) westward (cluster A), while the second included the

remainder of the specimens (cluster B) (Figure 3.1a). However, delta-K statistic might not be reliable under complex evolutionary scenarios (Janes et al. 2017). Additionally, while the PcoA also supported the same two main groups, with the first two components explaining 12% of the genetic variation (Figure 3.1d), results for increasing values of K indicated strong genetic structure across geographic space (Figure S3.1c). Therefore, we performed a hierarchical clustering analysis (run 2) for the two main clusters A and B (Figure 3.1b).

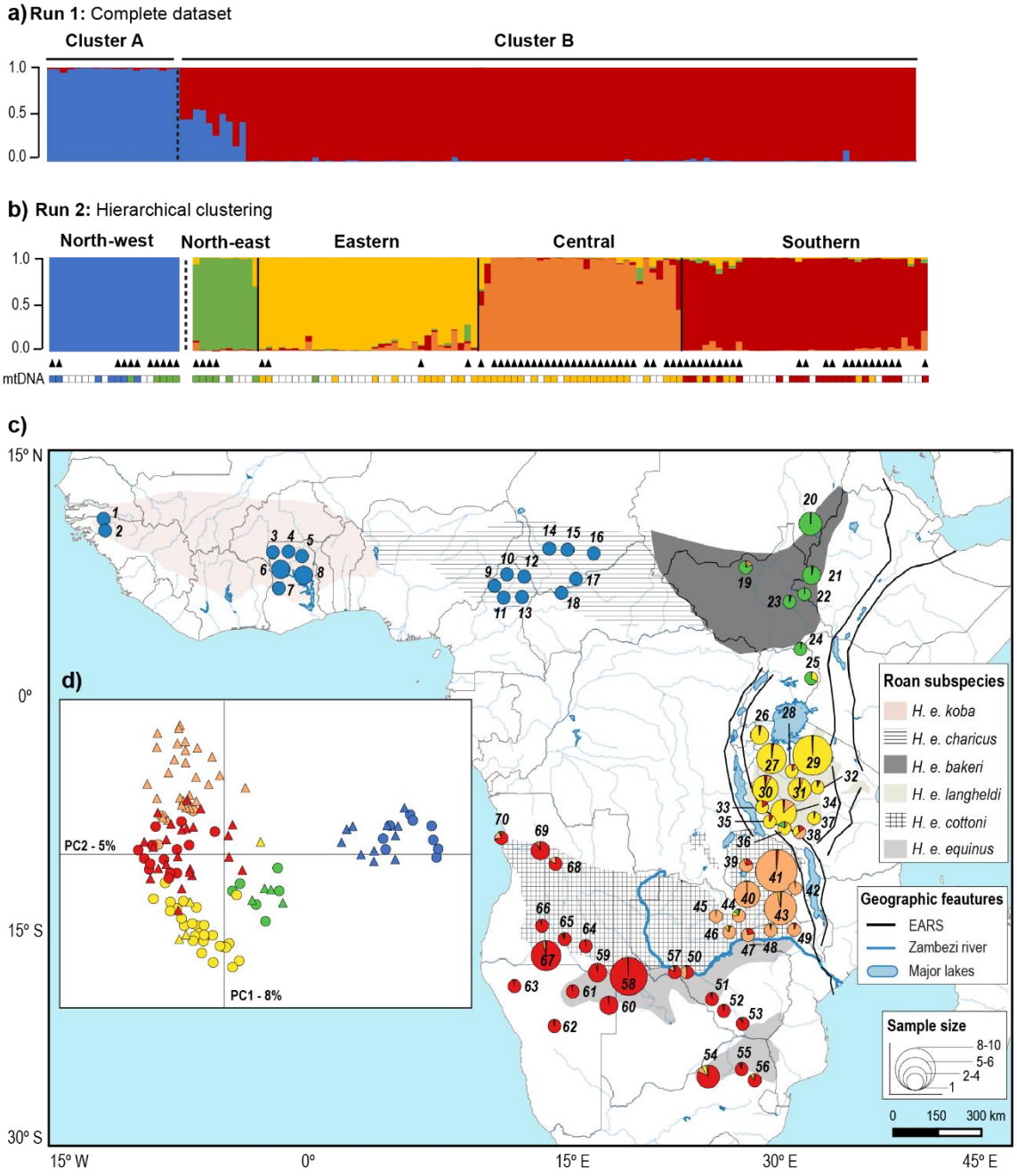


Figure 3. 1 Nuclear microsatellite data analyses for the roan antelope. **a)** Bayesian clustering analysis results for K = 2, using the complete dataset (run 1); **b)** Bayesian hierarchical clustering analyses results for K = 1 (cluster A) and K = 4 (cluster B). Each graph shows the representation of 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 main 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 S3.1. Triangles indicate historic samples. Also represented is the correspondent mitochondrial lineage for each individual: north-west in blue; north-east in green; east-central in yellow; and southern in red. Non-coloured triangles represent individuals with no sequenced mtDNA. **c)** Geographic distribution of nuclear groups, according to the Structure results from the hierarchical clustering analyses, and the five mains groups identified (Figure 3.1b). Roan antelope samples' distribution according to sampling location, indicated by numbers in italic black (from 1 to 70, as in Table S3.1). Pie-charts represent the average assignment probability to each cluster of individuals sampled within each location. Species distribution retrieved from IUCN (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 components from PcoA results, using the complete dataset (as in Figure 3.1a). Colours correspond to the main five groups identified from the hierarchical clustering analyses. Each point represents an individual. Circles depict contemporary samples, whereas triangles represent historic samples.

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 S3.2a), with subsequent values resulting in the split of individuals at the level of the sampling locality (Figure S3.2c). Results for Cluster B indicated K = 4 has the one with the highest likelihood (Figure S3.3a). The first group 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 3.1b and Figure S3.3c). Combining all results from the clustering analyses, five spatially explicit clusters were considered to best explain the SGS for the roan antelope (Figure 3.1c). These were designated, from west to east and south, as: north-west, north-east, eastern, central, and southern groups (Figure 3.1).

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 exclusively so, 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 3.1c – sampling site 19), on the border between Tanzania and Zambia between the eastern and central groups (Figure 3.1c – sampling site 38), and between the central and southern groups (Figure 3.1c – sampling site 50). For subsequent analyses, these admixed individuals were included in the groups for which they had the highest assignment probability.

The consistency of these five groups across time was tentatively (because of non-ideal sampling regimes for temporal data) assessed through a third Structure analysis considering contemporary and historical samples separately. For both datasets, the highest likelihood was assigned to K = 3 (Figure S3.4a and Figure S3.5a, respectively).

The clusters were mostly concordant with those derived from the hierarchical analyses from run 2 (Figure S3.4c and Figure S3.5c). Exceptions are in the assignment of the north-eastern individuals to a separate cluster, which only occurred for subsequent K = 4 for both sample types. Also, specimens from Zambia (the central group) did not fall out as a distinct group across contemporary samples, and the same can be observed for specimens from Tanzania (eastern group) across historical samples. We believe that low sample size for these regions (n = 3 Zambian individuals for contemporary samples – Figure S3.4; n = 4 Tanzanian individuals for historical samples – Figure S3.5), 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 predate 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.

3.3.1.2 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_{pa} statistics; Table 3.1). 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 S3.5). The southern group is characterized by a significant inbreeding coefficient (F_{IS}) for the combined dataset ($F_{IS} = 0.127$; $p < 0.001$; Table 3.1) as well as for historical samples ($F_{IS} = 0.165$; $p < 0.001$; Table S3.5), but not for the contemporary samples. This finding may simply reflect the substructure detected for the combined data (Figures S3.1c and S3.3c) and historical samples (Figure S3.5c), resulting in a deficit of heterozygotes; a well-known phenomenon known as the Wahlund effect (Waples 2015).

Table 3. 1 Measures of nuclear genetic diversity within each roan antelope group from the analysis of 43 microsatellites.

	N	H_e ± s.d.	N_a ± s.d.	N_{pa}	AR	F_{is}
North-west	20	0.601 ± 0.331	6.5 ± 3.9	52	5.0	0.110
North-east	10	0.580 ± 0.305	4.7 ± 2.7	12	4.5	0.043
Eastern	33	0.528 ± 0.330	6.0 ± 4.2	15	4.3	0.018
Central	31	0.503 ± 0.327	5.2 ± 3.5	9	3.8	0.070
Southern	37	0.521 ± 0.331	6.4 ± 4.7	26	4.3	0.127***

N, sample size; **H_e**, expected heterozygosity and respective standard deviations (**s.d**); **N_a**, average number of alleles per locus and respective standard deviations (**s.d**); **N_{pa}**, number of private alleles; **AR**, allelic richness; **F_{IS}**, inbreeding

coefficient with *** $p \leq 0.001$, after Bonferroni corrections. All parameters were estimated allowing 5% of missing data per locus.

Pairwise F_{ST} and G'_{ST} supported significant structure ($p < 0.001$) across the range of roan antelopes (see Table S3.6). 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 3.1d).

Our examination of population differentiation using analyses of molecular variance (AMOVA) found significant partitioning of variation between groups (10.5%), compared to 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 S3.7).

3.3.1.3 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 ten-fold difference. The fastest mutation rate ($1x10^{-3}$ 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 3.2). 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 3.2). 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 ($1x10^{-4}$ mutations/locus/year) for microsatellites resulted in a ten-fold increase in N_e and in the estimated intervals (Figure S3.6). This leads to a shift in the interpretation of events to older historical periods, but not on general population demographic trends.

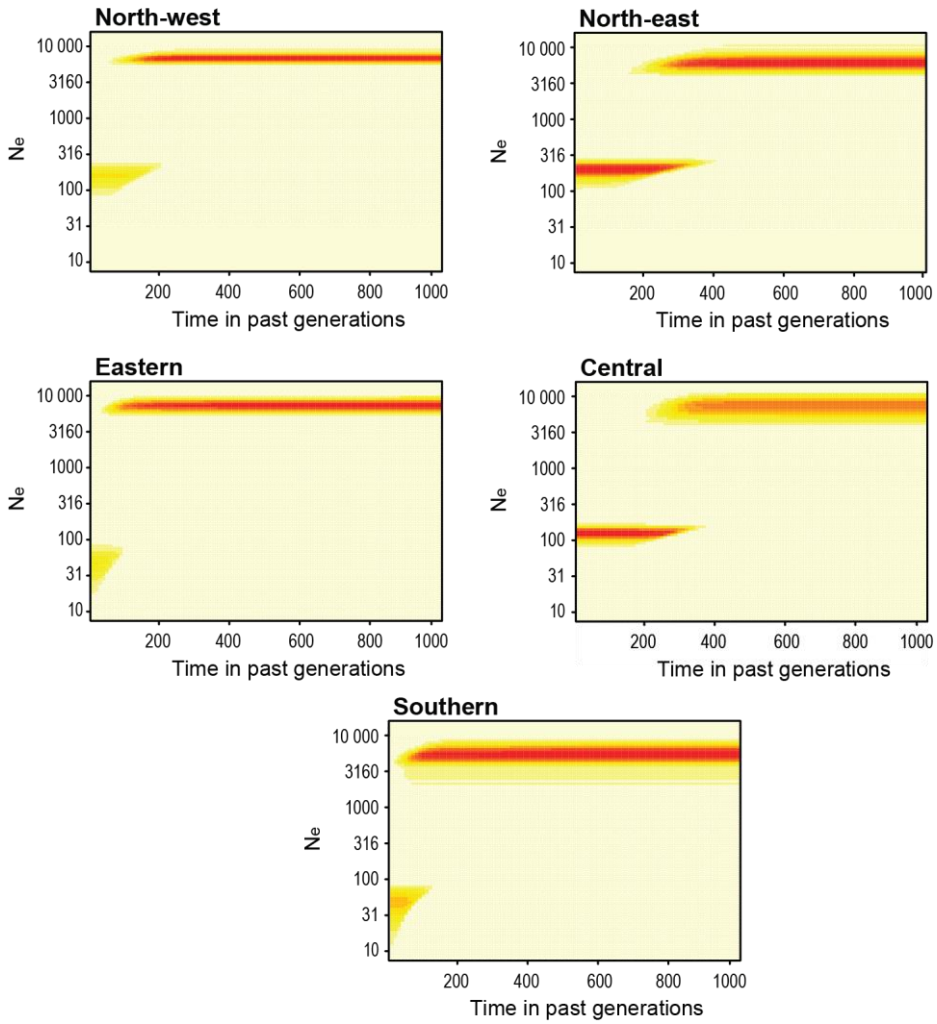


Figure 3. 2 Graphical two-dimensional summary of the posterior distribution of the estimated effective population size (N_e) for each nuclear group, using a mutation rate of $1x10^{-3}$ and two-step mutation model. The y-axis shows the N_e in logarithmic scale and x-axis shows the timescale in past generations, from present (generation 0) to 1,000 generation-time. Red colour represents maximum posterior values; yellow the minimum posterior values.

3.3.2 Mitochondrial data

For the mitogenomic data, our dataset comprised 15,551 bp (excluding the hypervariable region) for 82 roan antelope samples representative of the range of roan across Africa.

3.3.2.1 Diversity and differentiation

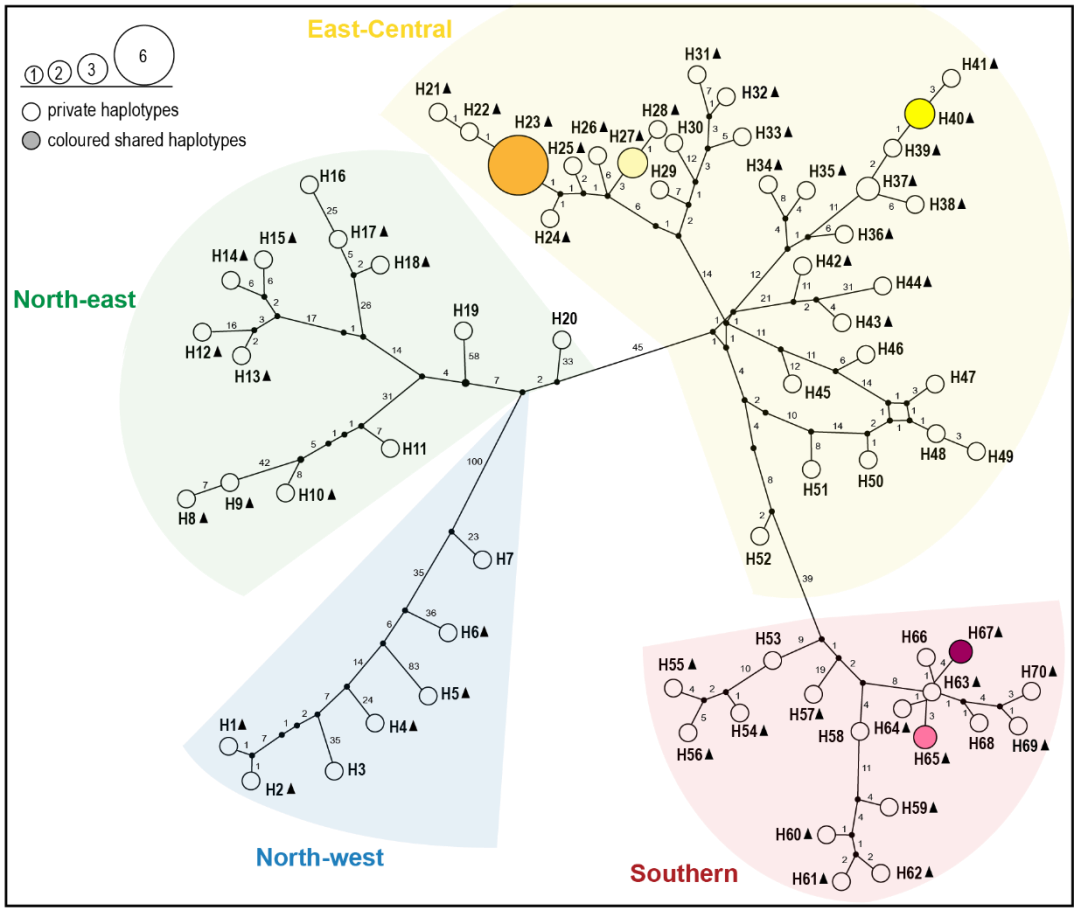
The average haplotype and nucleotide diversities across the mitogenome were 0.993 and 0.0059, respectively. Haplotype diversity ranged between 0.971 for cytochrome c oxidase subunit I (*cox1*) to 0.387 for the NADH dehydrogenase subunit 3 gene (*nad3*). For nucleotide diversity, values varied between π = 0.0109 in NADH dehydrogenase

subunit 4L (*nad4L*) and 0.0014 for the 12S rRNA subunit (Table S3.8). 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 3.3a). 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 3.1b and 3.3b). 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 3.3b – 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. While 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 3.3). Haplotypes from both lineages were present in the same localities in South Africa (Figure 3.3b – sampling site 54) and Angola (Figure 3.3b – sampling site 67). Also, in northern Tanzania, where the east-central lineage occurs, a few individuals had a mitochondrial haplotype more typical of the north-eastern lineage (Figure 3.3b – 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 3.3).

The north-west and north-east lineages have the highest values across all diversity estimates, including the number of substitution sites, and nucleotide and haplotype diversities (Table 3.2). A steep decrease in diversity estimates is seen from East Africa to Southern Africa, with the lowest values registered for the southern lineage (Table 3.2). 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 3.2). AMOVA, based on the mitochondrial data, indicated that 27% of the variation is explained by differences between the groups ($p < 0.001$) (Table S3.7). 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 S3.9).

a)


East-Central
North-east
North-west
Southern
private haplotypes
coloured shared haplotypes
H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 H12 H13 H14 H15 H16 H17 H18 H19 H20 H21 H22 H23 H24 H25 H26 H27 H28 H29 H30 H31 H32 H33 H34 H35 H36 H37 H38 H39 H40 H41 H42 H43 H44 H45 H46 H47 H48 H49 H50 H51 H52 H53 H54 H55 H56 H57 H58 H59 H60 H61 H62 H63 H64 H65 H66 H67 H68 H69 H70

b)

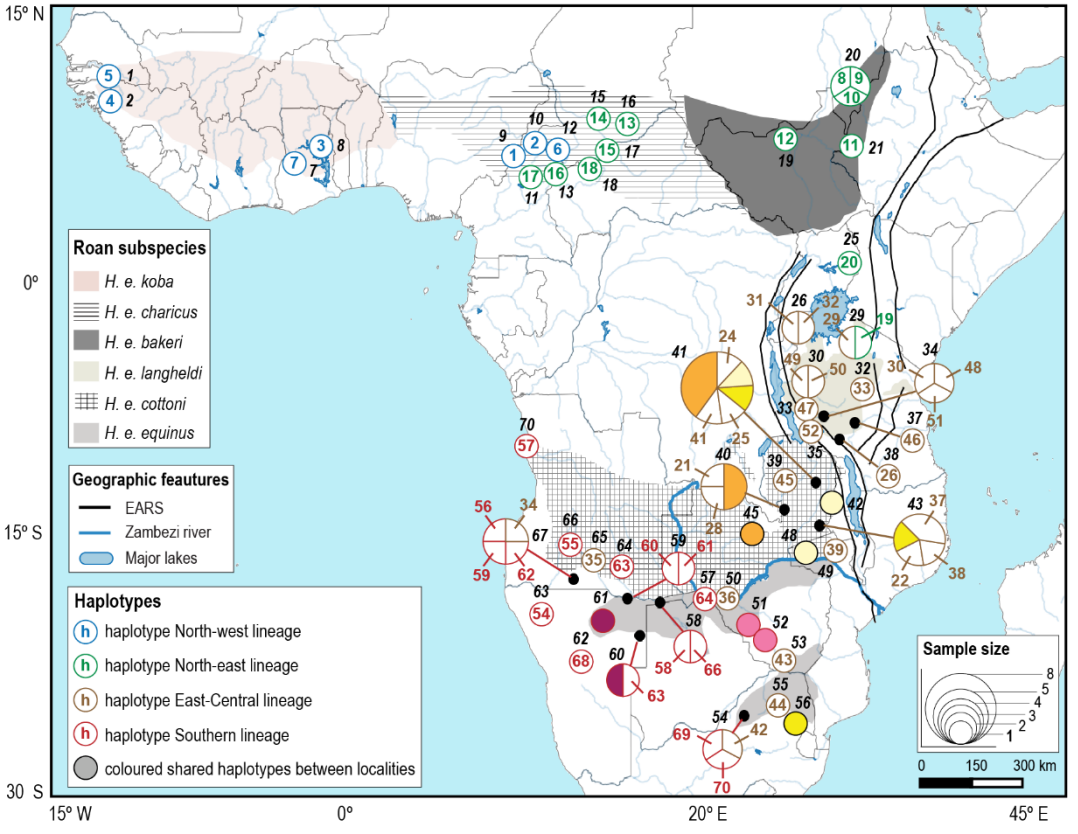

15° N
0°
15° S
30 S
15° W
0°
20° E
45° E
Roan subspecies
H. e. koba
H. e. charicus
H. e. bakeri
H. e. langheldi
H. e. cottoni
H. e. equinus
Geographic feautures
EARS
Zambezi river
Major lakes
Haplotypes
haplotype North-west lineage
haplotype North-east lineage
haplotype East-Central lineage
haplotype Southern lineage
coloured shared haplotypes between localities
Sample size
0 150 300 km

Figure 3. 3 Mitochondrial data analyses for the roan antelope. **a)** Median-Joining network based on mitochondrial genomes. Each lineage is identified by its name in bold and 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 corresponding number, preceded by an H. Triangles next to haplotype name indicate those that are present in historic samples. Size of the circle is scaled according to the number of individuals presenting that haplotype. Numbers in bold on top of the branches represent the number of mutation steps between haplotypes and lineages. Missing haplotypes are represented by small black circles. Private haplotypes are represented in white, whereas shared haplotypes between sampling localities are coloured. **b)** Geographical representation of the mitochondrial haplotypic network. Each circle characterizes the haplotypic composition of individuals within each sampling locality. Haplotypes contoured according to each lineage: north-west in blue; north-east in green; east-central in brown; and southern in red. Shared haplotypes between localities indicated by full coloured circles, according to Figure 3a. Roan antelope samples' distribution according to sampling location, indicated by numbers in italic black (from 1 to 70, as in Table S3.1). Species distribution retrieved from IUCN (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.

Table 3. 2 Measures of genetic diversity within each roan mitochondrial lineage (top) and nuclear group (bottom), based on whole mitochondrial DNA sequences.

MITOCHONDRIAL LINEAGES

	N	S	h	H_d (± s.d.)	π (± s.d.)	MPD (± s.d.)
North-west	7	246	7	1.000 ± 0.076	0.00540 ± 0.00304	83.810 ± 41.224
North-east	13	256	13	1.000 ± 0.030	0.00455 ± 0.00236	70.718 ± 32.632
East-Central	40	221	32	0.974 ± 0.015	0.00238 ± 0.00118	35.073 ± 16.459
Southern	22	95	18	0.994 ± 0.019	0.00146 ± 0.00075	22.661 ± 10.443

NUCLEAR GROUPS

	N	S	h	H_d (± s.d.)	π (± s.d.)	MPD (± s.d.)
North-west	13	432	13	1.000 ± 0.030	0.00832 ± 0.00429	129.308 ± 59.390
North-east	6	319	6	1.000 ± 0.096	0.00918 ± 0.00533	142.800 ± 71.749
Eastern	13	254	13	1.000 ± 0.030	0.00421 ± 0.00219	65.462 ± 30.231
Central	23	202	14	0.921 ± 0.042	0.00275 ± 0.00138	42.751 ± 19.253
Southern	27	229	25	0.994 ± 0.012	0.00369 ± 0.00183	57.325 ± 25.557

N, number of samples; **S**, number of polymorphic (segregating) sites; **h**, number of haplotypes; $\mathbf{H_d}$, haplotype diversity and respective standard deviation (s.d.); **π**, nucleotide diversity and respective standard deviation (s.d.); **MPD**, mean pairwise distance and respective standard deviation (s.d.). All parameters were estimated allowing 5% of missing data.

3.3.2.2 Divergence time estimates and demography

The inferred rate of change for the three partitions analysed varied between $2.3x10^{-8}$ mutations/site/year for the 3rd codon-position of the protein-coding genes (95% HPD: $2.1–2.5x10^{-8}$) to $4.2x10^{-9}$ (95% HPD: $3.6–4.7x10^{-9}$) and $5.1x10^{-9}$ (95% HPD: $4.8–5.8x10^{-9}$) mutations/site/year for the 1st and 2nd 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 3.4). 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 *Hippotragus* spp. 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).

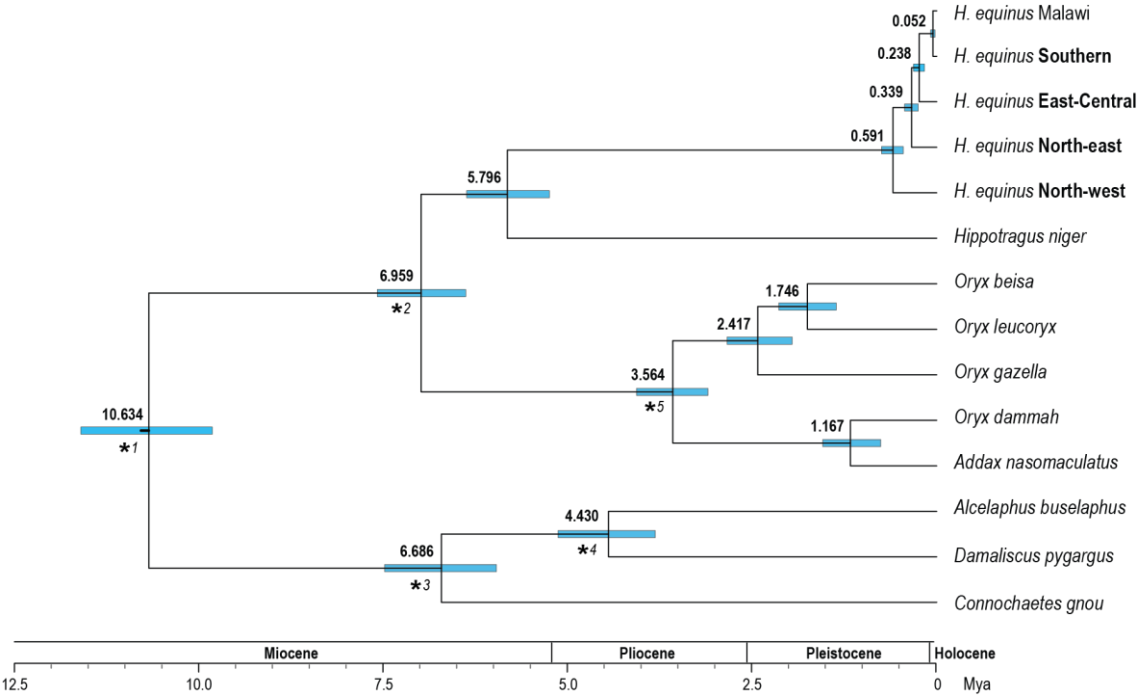


Figure 3. 4 Bayesian phylogenetic analysis of 10 reference mitogenomes (excluding hyper-variable region - HVR) within Hippotraginae and Alcelaphinae sub-families (species names in italic – see text for more info on available reference) and each of the four roan mitochondrial lineages (names in bold - represented by a single sample). All nodes had a posterior probability of 1.0. The 95% higher posterior density (HPD) intervals for divergence time estimates (blue bars) are indicated for each node, with corresponding mean value (in black), according to timescale at the bottom, in million years ago (Mya). Calibrated nodes, following Bibi (2013), are indicated with an asterisk, and identified by a number. Check Table S3.3 for details on prior mean distribution and standard deviation.

Neutrality tests across the mitochondrial data, retrieved negative albeit non-significant values across all lineages (Table S3.10), indicating no departure from the null hypothesis of demographic stability.

3.4 Discussion

Roan antelopes have a wide distribution across Sub-Saharan African savannas. By combining mitogenomes with nuclear genomic information of a suite of 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, past climatic events, and geomorphological features that shaped African biodiversity during the Pleistocene.

3.4.1 Genetic structure and congruence with subspecies

Based on geographical range, Ansell (1972) accepted and recognized six subspecies for the roan antelope. The validity and geographical correspondence of these subspecies have been questioned by genetic data (Matthee and Robinson 1999; Alpers et al. 2004) and are not currently recognized by the IUCN (East and IUCN SSC Antelope Specialist Group 1999; IUCN SSC Antelope Specialist Group 2017). Nevertheless, we found near-perfect congruence between our distinct genetic groups and these six subspecies. For both nuclear and mitochondrial data, the most notable difference includes *H. e. charicus*, the subspecies originally described from the area between Benin and Sudan. Rather, individuals from this region belong to our 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. Remaining differences between the range of genetic groups and subspecies are observed at the nuclear level for individuals sampled in southern Uganda, which belong to *H. e. bakeri* (our north-east group) rather than to *H. e. langheldi* (our eastern group), and for Angolan samples, which are genetically similar to individuals of the nominal *equinus* subspecies rather than to *H. e. cottoni* subspecies. At the mitochondrial level, the main difference relates to the east-central lineage that encompasses the range of both *H. e. langheldi* and *H. e. cottoni*. Some individuals in Southern Africa, as far as Angola, also share haplotypes from this east-central lineage, besides the recognized *H. e. equinus* (our southern lineage). To the best of our knowledge, this is the first comparison of Angolan roan antelope mtDNA with other regions. Two sites to the north and east of Lake Victoria have mitotypes typical of *H. e. bakeri* (north-east lineage), as was also highlighted by Alpers et al. (2004). Interestingly, both Matthee and Robinson (1999) and Alpers et al. (2004) have previously reported a

well differentiated northwest mtDNA lineage, as well as a unique lineage for *H. e. langheldi* and *H. e. cottoni* subspecies, although Matthee and Robinson (1999) classify them as the same subspecies in opposition to Ansell's work. Additionally, due to shortage sampling covering the specie's range, Alpers et al. (2004) were unable to assign samples from Cameroon to any mitochondrial lineage, and left it unresolved. It is possible that the observed differences in the present study between the mitogenome and nuclear signatures may have been driven by male-biased dispersal in the species (Chardonnet and Crosmary 2013), given the maternal mode of inheritance of mitochondrial DNA.

3.4.2 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 (McDonough et al. 2015; Bertola et al. 2016; Visser, Bennett and Jansen van Vuuren 2019). The Tropical Forest Belt spans from West Africa eastward to the East African Rift System, and has been invoked as a major barrier for savanna-adapted terrestrial species. Large ungulates like hartebeest and the common 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 to 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 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 East African Rift System (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 in large mammals, including in ungulates such as waterbuck (Lorenzen et al. 2006), giraffe (Lorenzen et al. 2012; Winter, Fennessy and Janke 2018) or sable antelope (Jansen van Vuuren et al. 2010; Vaz Pinto 2018), but also for carnivores including African lion and cheetah (Charruau et al. 2011; Bertola et al. 2016). 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, the Zambezi River is also a major driver of differentiation within mammals in Southern Africa (e.g. McDonough et al. 2015; Vaz Pinto 2018). Extreme hydrographic changes that took place during the mid-Pleistocene in Southern Africa (Walford, White and Sydow 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 hybridization) zones in many species. The contact zone between our north-east and eastern roan 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 in the roan 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 elephant is recent and probably the result of habitat modifications and poaching. In roan antelopes, this contact zone could have been formed as a result of north-east mitochondrial introgression into the north-west group, or by predominantly north-west male dispersal towards the north-east group.

3.4.3 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-west towards the east and southern groups, supporting a Pleistocene refugia in West Africa (previously proposed by Alpers et al. 2004); we also speculate that West African roan antelopes are the ancestral population based on divergence dates between groups. Our findings follow the classical scenario where repeated colonisations of new areas through range expansions, systematically reduce genetic diversity, resulting in a typical population structure (Excoffier and Ray 2008; Excoffier, Fool and Petit 2009). A West African origin for taxonomic groups has similarly been proposed for other ungulates including hartebeest, buffalo, and giraffe (Flagstad et al. 2001; Brown et al. 2007; 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 decrease possibly triggered by severe Holocene droughts together with increased human-pressure at the beginning of the Neolithic, as happened for other water-dependant ungulate species such as the African buffalo (Heller et al. 2008; Heller, Brüniche-Olsen and Siegismund 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 and IUCN SSC Antelope Specialist Group 1999).

However, the species is virtually extinct in many regions, 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 e.g. Jansen van Vuuren, Rushworth and Montgelard 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 and 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 and André 2013). For species subjected to common climatic and geological events, such as African ungulates, the location of refugial areas

and the presence of geomorphological barriers, dictated subsequent range expansion rates, and currently observed spatial patterns (Lorenzen et al. 2012). Shared evolutionary histories can dictate common responses to current threats and, therefore, the continued existence of a large number of species.

3.5 References

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Ansell, W. F. H., 1972. Order Artiodactyla, pp. 15–83 in *Mammals of Africa: an Identification Manual*. Smithsonian Institute Press (J. Meester and H. W. Setzer, Eds.), Washington, DC.

Arctander, P., Johansen, C. and Coutellec-Vreto, M.-A., 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution*, 16: 1724–1739.

Badgley, C., 2010. Tectonics, topography, and mammalian diversity. *Ecography*, 33: 220–231.

Bandelt, H., Forster, P. and Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16: 37–48.

Bertola, L. D., Jongbloed, H., Van Der Gaag, K. J., de Kniijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., et al., 2016. Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the Lion (*Panthera leo*). *Scientific Reports*, 6: 30807.

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

Bolger, A. M., Lohse, M., & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30: 2114–2120.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A. and Drummond, A. J., 2014. BEAST 2: a software platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10: e1003537.

Brown, D. M., Brenneman, R. A., Koepfli, K.-P., Pollinger, J. P., Milá, B., Georgiadis, N. J., Louis, E. E., Grether, G. F., Jacobs, D. K. and Wayne, R. K., 2007. Extensive population genetic structure in the giraffe. *BMC Biology*, 5: 57-70.

Bryant, D. and Moulton, V., 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21: 255–

265.

Chardonnet, P. and Crosmary W., 2013. *Hippotragus equinus* Roan Antelope, pp. 548–556 in *Mammals of Africa* VI. Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Charruau, P., Fernandes, C., Orozco-Terwengel, P., Peters, J., Hunter, L., Ziaie, H., Jourabchian, A., Jowkar, H., et al., 2011. Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. *Molecular Ecology,* 20: 706-724.

Chorowicz, J., 2005. The East African rift system. *Journal of African Earth Sciences*, 43: 379–410.

Collen, B., Turvey, S. T., Waterman, C., Meredith, H. M., Kuhn, T. S., Baillie, J. E. and Isaac, N. J., 2011. Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366: 2611–2622.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, 52: 87-94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763.

DeMenocal, P. B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, 220: 3–24.

Dupont, L., 2011. Orbital scale vegetation change in Africa. *Quaternary Science Reviews*, 30: 3589–3602.

Du Toit, J. T. and Cumming, D. H. M., 1999. Functional significance of ungulate diversity in African savannas and the ecological implications of the spread of pastoralism. *Biodiversity and Conservation*, 8: 1643–1661.

Earl, D. A. and VonHoldt, B. M., 2012. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources*, 4: 359–361.

East, R. and IUCN SSC Antelope Specialist Group, 1999. African antelope database 1998 (IUCN/SSC Antelope Specialist Group, Ed.), Gland: IUCN.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32: 1792–1797.

Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, 14:

2611–2620.

Excoffier, L. and Ray, N., 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution,* 23: 347-351.

Excoffier, L., Foll, M. and Petit, R. J., 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics,* 40: 481-501.

Excoffier, L. and Lischer., H. E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564–567.

Falush, D., Stephens, M. and Pritchard, J. K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics Society of America*, 164: 1567–1587.

Flagstad, Ø., Syversten, P. O., Stenseth, N. C. and Jakobsen, K. S., 2001. Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climatic variation. *Proceedings of the Royal Society of London B: Biological Sciences*, 268: 667–677.

Fu, Y.-X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147: 915–925.

Goudet, J., 2001. Fstat, version 2.9.3.2. A program to estimate and test gene diversities and fixation indices. Institute of Ecology University of Lausanne, Switzerland.

Goudie, A. S., 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67: 437–456.

Grobler, J. P., Pretorius, D. M., Botha, K., Kotze, A., Hallerman, E. M. and Jansen van Vuuren, B., 2005. An exploratory analysis of geographic genetic variation in southern African nyala (*Tragelaphus angasii*). *Mammalian Biology*, 70: 291–299.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59: 307–321.

Hartmann, K. and André, J., 2013. Should evolutionary history guide conservation? *Biodiversity and Conservation*, 22: 449–458.

Hedrick, P. W., 2005. A standardized genetic differentiation measure. *Evolution*, 59: 1633–1638.

Heller, R., Lorenzen, E. D., Okello, J. B. A., Masembe, C. and Siegismund, H. R., 2008. Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Molecular Ecology,* 17*:* 4845-4858.

Heller, R., Brüniche-Olsen, A. and Siegismund, H. R., 2012. Cape buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. *Molecular*

Ecology, 21: 3947-3959.

Hewitt, G., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359: 183–195.

Hubisz, M. J., Falush, D., Stephens, M. and Pritchard, J. K., 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9: 1322–1332.

IUCN SSC Antelope Specialist Group, 2017. *Hippotragus equinus*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10167A50188287.en on April 8th, 2020.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I. and Andrew, R. L., 2017. The K = 2 conundrum. *Molecular Ecology*, 26: 3594–3602.

Jansen van Vuuren, B., Rushworth, I. and Montgelard, C., 2017. Phylogeography of oribi antelope in South Africa: evolutionary versus anthropogenic panmixia. *African Zoology*, 52: 189–197.

Kingdon, J., 2013. Mammalian evolution in Africa, pp. 75–100 in *Mammals of Africa I*. Bloomsbury Publishing (J. Kingdon, D. Happold, M. Hoffmann, T. Butynski, M. Happold and J. Kalina, Eds.), London.

Kircher, M., Sawyer, S. and Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40: e3–e3.

Lanfear, R., Calcott, B., Ho, S. Y. and Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29: 1695–1701.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34: 772–773.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25: 2078–2079.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*: 1303–3997.

Librado, P. and Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25: 1451–1452.

Lorenzen, E. D., Simonsen, B. T., Kat, P. W., Arctander, P. and Siegismund, H. R., 2006.

Hybridization between subspecies of waterbuck (*Kobus ellipsiprymnus*) in zones of overlap with limited introgression. *Molecular Ecology*, 15: 3787–3799.

Lorenzen, E. D., De Neergaard, R., Arctander, P. and Siegismund, H. R., 2007. Phylogeography, hybridization and Pleistocene refugia of the kob antelope (*Kobus kob*). *Molecular Ecology*, 16: 3241–3252.

Lorenzen, E. D., Masembe, C., Arctander, P. and Siegismund, H. R., 2010. A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: insights from mtDNA and the common eland antelope. *Journal of Biogeography*, 37: 571–581.

Lorenzen, E. D., Heller, R. and Siegismund, H. R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21: 3656–3670.

Matthee, C. A. and Robinson, T. J., 1999. Mitochondrial DNA population structure of roan and sable antelope; implications for the translocation and conservation of the species. *Molecular Ecology*, 8: 227–238.

Mayaux, P., Bartholomé, E., Fritz, S. and Belward, A., 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31: 861–877.

McDonough, M. M., Šumbera, R., Mazoch, V., Ferguson, A. W., Phillips, C. D. and Bryja, J., 2015. Multilocus phylogeography of a widespread savanna–woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology*, 24: 5248–5266.

Meyer, M. and Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010: pdb.prot5448.

Mondol, S., Moltke, I., Hart, J., Keigwin, M., Brown, L., Stephens, M. and Wasser, S. K., 2015. New evidence for hybrid zones of forest and savanna elephants in Central and West Africa. *Molecular Ecology*, 24: 6134–6147.

Moorley, R. and Kingdon, J., 2013. Africa's environmental and climatic past, pp. 43–56 in *Mammals of Africa I*. Bloomsbury Publishing (K. J, D. Happold, M. Hoffmann, T. Butynski, M. Happold and J. Kalina, Eds.), London.

Muwanika, V. B., Nyakaana, S., Siegismund, H. R. and Arctander, P., 2003. Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, 91: 361–372.

Nikolic, N. and Chevalet, C., 2014. Detecting past changes of effective population size. *Evolutionary Applications*, 7: 663–681.

O'Connell, K. A., Mulder, K. P., Maldonado, J., Currie, K. L. and Ferraro, D. M., 2019. Sampling related individuals within ponds biases estimates of population structure

in a pond-breeding amphibian. *Ecology and Evolution*, 9: 3620–3636.

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Grottolo Marasini, G., Visconti, P. and Rondinini, C., 2013. Generation length for mammals. *Nature Conservation,* 5: 87–94.

Peakall, R. and Smouse, P. E., 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6: 288–295.

Peakall, R. and Smouse, P. E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28: 2537–2539.

Pritchard, J. K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.

Pritchard, J. K., Wen, X. and Falush, D., 2009. Documentation for Structure software: version 2.3. University of Chicago, Chicago. Retrieved from http://web.stanford.edu.

Qi, W.-H., Jiang, X.-M., Du, L.-M., Xiao, G.-S., Hu, T.-Z., Yue, B.-S., Quan, Q.-M. and Munderloh, U. G., 2015. Genome-wide survey and analysis of microsatellite sequences in bovid species. *Plos One*, 10: e0133667.

Rambaut, A., Drummond, A., Xie, D., Baele, G. and Suchard, M. A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67: 901–904.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Rice, W. R., 1989. Analysing tables of statistical tests. *Evolution*, 43: 223–225.

Ritland, K., 1996. Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research*, 67: 175–185.

Ritz, L. R., Glowatzki-Mullis, M., MacHugh, D. E. and Gaillard, C., 2000. Phylogenetic analysis of the tribe Bovini using microsatellites. *Animal Genetics*, 31: 178–185.

Rocha, J., 2014. The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences. MsC thesis dissertation, University of Porto, Portugal.

Scott, P. A., Allison, L. J., Field, K. J., Averill-Murray, R. C. and Shaffer, H. B., 2020. Individual heterozygosity predicts translocation success in threatened desert tortoises. *Science,* 370: 1086-1089.

Shettima, B., Kyari, A., Aji, M. and Adams, F., 2018. Storm and tide influenced depositional architecture of the Pliocene–Pleistocene Chad Formation, Chad Basin (Bornu Sub–basin) NE Nigeria: a mixed fluvial. *Journal of African Earth Sciences*, 143: 309–320.

Simpson, J., Wong, K., Jackman, S., Schein, J., Jones, S. J. and Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research*, 19: 1117-1123.

Smitz, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., Caron, A., Prins, et al., 2013. Pan-African genetic structure in the African Buffalo (*Syncerus caffer*): investigating intraspecific divergence. *Plos One*, 8: e56235.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics Society of America*, 123: 585–595.

Tan, J. and Mikheyev, A., 2016. A scaled-down workflow for Illumina shotgun sequencing library preparation: lower input and improved performance at small fraction of the cost. *PeerJ*, 4: e2475v1.

Trauth, M. H., Larrasoaña, J. C. and Mudelsee, M., 2009. Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, 28: 399–411.

Vaz Pinto, P., Lopes, S., Mourão, S., Baptista, S., Siegismund, H. R., Jansen van Vuuren, B., Beja, P., Ferrand, N. and Godinho, R., 2015. First estimates of genetic diversity for the highly endangered giant sable antelope using a set of 57 microsatellites. *European Journal of Wildlife Research*, 61: 313–317.

Vaz Pinto, P., 2018. Evolutionary history of the critically endangered giant sable antelope (*Hippotragus niger variani*). Insights into its phylogeography, population genetics, demography and conservation. PhD thesis dissertation, University of Porto, Portugal.

Visser, J. H., Bennett, N. C. and Jansen van Vuuren, B., 2019. Phylogeny and biogeography of the African Bathyergidae: a review of patterns and processes. *PeerJ*, 7: e7730.

Vrba, E., 1996. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate, pp. 385–424 in *Paleoclimate and Evolution*. Yale University Press (E. Vrba, G. Denton, T. Partidge and L. Burckle, Eds.), New Haven, Connecticut.

Walford, H. L., White, N. J. and Sydow, J. C., 2005. Solid sediment load history of the Zambezi Delta. *Earth and Planetary Science Letters*, 238: 49–63.

Waples, R. S., 2015. Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity,* 106: 1-19.

Winter, S., Fennessy, J. and Janke, A., 2018. Limited introgression supports division of giraffe into four species. *Ecology and Evolution*, 8: 10156–10165.

3.6 Supporting Information

Text S3.1 Sample collection and DNA extraction. Contemporary data comprised muscle samples, collected from 1993 to 2004 and donated by several individuals (Table S3.1). DNA extraction from modern tissue samples was done using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germany), following the manufacturer's instructions. Historic samples were donated by five Museum collections and comprised dry skin, skull, bone, and teeth material from samples collected from 1842 to 1973 (Table S3.1). All laboratory procedures for historic samples were performed on a dedicated laboratory for low quality DNA, with high-pressure chamber, ultra-violet sterilization of buffer solutions and remaining equipment, as well as usage of personal protection gear. DNA extraction followed Dabney et al. (2013) protocol. Dry skin samples were re-hydrated. Sampled material was previously digested for 24h to 48h in a hoven at 32º C, together with a digestion buffer containing proteinase K. Teeth powder for digestion was obtained from the pulp, after drilling through the dentin, using a Dremel diamond drill, sterilized in between each utilization. Bone powder for digestion was obtained by pulverization in a bone mill. Purification and elution steps were performed using MinElute PCR purification kit (Qiagen, Germany). For each historical sample, we performed triplicates of extraction to guarantee a final eluate of 150 ul of purified DNA. Geographic sample location followed sampling information. For historical samples, this information was complemented by searching journals of African explorations, given the author and year of collection.

Text S3.2 Microsatellite amplification. In order to increase amplification success, the following adjustments were done to the protocol followed as in Vaz Pinto et al. (2015): 1) multiplex reactions were split from the original ones, reducing the number of loci per reaction (Table S3.2); 2) each multiplex reaction was perform in two rounds – a first one with a touchdown in annealing temperature and a second one with lower annealing temperature and more cycles (Table S3.2); 3) a pre (without fluorescence dye) and post-PCR reaction (with reverse only primer, fluorescent dye and more cycles) was employed following Piggott and co-workers (2004)[1].

[1]Piggott, M. P., Bellemain, E., Taberlet, P. and Taylor, A. C., 2004. A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics*, 5: 417–420.

Text S3.3 mtDNA bait production. Primers used for the amplification of the four long-range PCR products, encircling the complete mitochondrial genome of roan antelope (*H.*

equinus) were the same as used in Rocha (2004) for sable (*H. niger*). However, thermal-cycler programs for primer-pairs P1 (Hn1377F and Hn6895R) and P2 (Hn6538F and Hn12024R) were slightly changed to lower specificity and increase annealing success, considering they amplified the longest fragments. For primer-pair P1, the thermal-cycling conditions were an initial denaturation at 92ºC for 2 min, followed by 30 cycles of a denaturation step at 92ºC for 20s, annealing at 60ºC for 45s and extension step at 68ºC for 6 min, followed by a final extension of 68ºC at 6 min. The same cycling conditions were applied to primer-pair P2, except 37 cycles were used after initial denaturation and annealing time on each cycle was of 20s.

Table S3. 1 Sampling for *Hippotragus equinus*, ordered by sampling site, country, locality gender, and type of material. The sampling site number used here is included in all the figures throughout the manuscript.

Sampling site	Sample	Country	Locality	Gender	Material	Donated by	Original label	Year	microsatellite dataset	mtDNA dataset
1	HeNI217	Guinea Bissau	Kopulan Camp	F	Skin	P-C Museum	GUINEA.6	1911	✓	✓
2	HeNI219	Guinea Bissau	Crocoli Camp	F	Skin	P-C Museum	GUINEA.31	1911	✓	✓
3	He79	Ghana	Jang	M	Muscle	H. R. Siegismund	7222	1998	✓	
4	He51	Ghana	Mole Nat. Park	—	Muscle	H. R. Siegismund	3852	1997	✓	
5	He68	Ghana	Brugbani Camp	M	Muscle	H. R. Siegismund	5544	1997	✓	
6	He85	Ghana	Kananto	M	Muscle	H. R. Siegismund	7506	1998	✓	
6	He86	Ghana	Kananto	F	Muscle	H. R. Siegismund	7507	1998	✓	
7	He80	Ghana	Kablima	F	Muscle	H. R. Siegismund	7324	1998	✓	✓a)
8	He87	Ghana	Kabanpe	—	Muscle	H. R. Siegismund	7510	1998	✓	
8	He89	Ghana	Kabanpe	M	Muscle	H. R. Siegismund	7512	1998	✓	✓
9	HeNI236	Cameroon	Faro River	F	Teeth	NHM Wien	NMW 4737	1933	✓	✓
10	HeNI240	Cameroon	Karba Petel	M	Skull	NHM Wien	NMW 5323	1937	✓	✓
11	HeNI227	Cameroon	N'Gaoundere	M	Skin	P-C Museum	CAMII.88	1931	✓	✓
12	HeNI229	Cameroon	Bi-Indu	M	Skin	P-C Museum	CAMII.117	1931	✓	✓
13	He59	Cameroon	Cameroon	M	Muscle	H. R. Siegismund	4763	1997	✓	✓
14	HeNI237	Chad	Schari River	M	Skull	NHM Wien	NMW 5288/B 5046	1937	✓	
15	HeNI220	Chad	Melfi	F	Skin	P-C Museum	NN.98	1925	✓	✓
16	HeNI222	Chad	Iro Lake	M	Skin	P-C Museum	NN.158	1925	✓	✓
17	HeNI224	Chad	Dalmija	M	Skin	P-C Museum	NN.218	1925	✓	✓
18	HeNI226	CAR	Nana Barya Res.	M	Skin	P-C Museum	NN.252	1925	✓	✓
19	HeNI216	South Sudan	Bahr el Ghazal	M	Skin	P-C Museum	SUDANI.106	1933	✓	✓
20	HeNI233	Sudan	Dinder	—	Teeth	NHM Wien	NMW 62040	1969	✓	✓
20	HeNI242	Sudan	Dinder	—	Teeth	NHM Wien	NMW 6015	1925	✓	✓
20	HeNI244	Sudan	Dinder	F	Skull	NHM Wien	NMW 6032	1925	✓	✓

21	He110	Ethiopia	Gambela Nat.Park	F	Muscle	H. R. Siegismund	9295	—	✓	
21	He111	Ethiopia	Gambela Nat. Park	M	Muscle	H. R. Siegismund	9296	—	✓	✓
22	He108	Ethiopia	Akobe septum	M	Muscle	H. R. Siegismund	9293	—	✓	
23	He112	South Sudan	Boma Nat. Park	F	Muscle	H. R. Siegismund	9486	2004	✓	
24	He29	Uganda	Kidepo Valley Nat. Park	—	Muscle	H. R. Siegismund	1234	1994	✓	
25	He105	Uganda	Pian Upe Game Res.	F	Muscle	H. R. Siegismund	9056	2000	✓	✓b)
26	HeNI182	Tanzania	Mawale (Kagera)	—	Bone	NHM Berlin	ZMB_Mam_067994	—	✓	✓
26	HeNI202	Tanzania	Mawale (Kagera)	—	Bone	NHM Berlin	ZMB_Mam_068210	—	✓	✓
27	He101	Tanzania	Kigosi Game Res.	M	Muscle	H. R. Siegismund	8633	1998	✓	
27	He57	Tanzania	Kigosi Game Res.	M	Muscle	H. R. Siegismund	4129	1996	✓	
27	He97	Tanzania	Kigosi Game Res.	M	Muscle	H. R. Siegismund	8252	1997	✓	
27	He98	Tanzania	Kigosi Game Res.	M	Muscle	H. R. Siegismund	8253	1997	✓	
27	He99	Tanzania	Kigosi Game Res.	M	Muscle	H. R. Siegismund	8631	1998	✓	
28	He63	Tanzania	Wembere	M	Muscle	H. R. Siegismund	5541	1997	✓	
29	He100	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	8632	1998	✓	✓
29	He103	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	8635	1998	✓	
29	He48	Tanzania	Maswa Game Res.	—	Muscle	H. R. Siegismund	3612	1994	✓	
29	He49	Tanzania	Maswa Game Res.	—	Muscle	H. R. Siegismund	3613	1994	✓	✓
29	He50	Tanzania	Maswa Game Res.	—	Muscle	H. R. Siegismund	3614	1993	✓	
29	He61	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	5534	1997	✓	
29	He62	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	5535	1997	✓	
29	He94	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	8249	1997	✓	
29	He95	Tanzania	Maswa Game Res.	M	Muscle	H. R. Siegismund	8250	1997	✓	
30	He102	Tanzania	Ugalla River Game Res.	M	Muscle	H. R. Siegismund	8634	1998	✓	✓
30	He104	Tanzania	Ugalla River Game Res.	M	Muscle	H. R. Siegismund	8636	1998	✓	
30	He65	Tanzania	Ugalla River Game Res.	M	Muscle	H. R. Siegismund	5543	1997	✓	

30	He66	Tanzania	Ugalla River Game Res.	M	Muscle	H. R. Siegismund	6462	1997	✓	✓c)
31	He45	Tanzania	Kizigo Game Res.	M	Muscle	H. R. Siegismund	3300	1995	✓	
31	He46	Tanzania	Kizigo Game Res.	—	Muscle	H. R. Siegismund	3610	1993	✓	
31	He47	Tanzania	Kizigo Game Res.	—	Muscle	H. R. Siegismund	3611	1993	✓	
32	HeNI187	Tanzania	Turu, Singida	M	Bone	NHM Berlin	ZMB_Mam_068017	—	✓	✓
33	He107	Tanzania	Katavi Nat. Park	M	Muscle	H. R. Siegismund	9111	2000	✓	✓
34	He44	Tanzania	Rungwa Game Res.	—	Muscle	H. R. Siegismund	3299	1995	✓	✓
34	He52	Tanzania	Rungwa Game Res.	—	Muscle	H. R. Siegismund	4121	—	✓	✓
34	He53	Tanzania	Rungwa Game Res.	M	Muscle	H. R. Siegismund	4123	1996	✓	
34	He54	Tanzania	Rungwa Game Res.	M	Muscle	H. R. Siegismund	4124	1996	✓	✓
35	He55	Tanzania	Mlele, Katavi	M	Muscle	H. R. Siegismund	4127	—	✓	✓
36	HeNI189	Tanzania	Lupa River	—	Bone	NHM Berlin	ZMB_Mam_068024	—	✓	
37	HeNI178	Tanzania	Unika	—	Bone	NHM Berlin	ZMB_Mam_067986	1900	✓	✓
38	He106	Tanzania	Ruaha Nat. Park	M	Muscle	H. R. Siegismund	9071	2000	✓	✓
39	He39	Zambia	Bangweulu	—	Muscle	H. R. Siegismund	2532	1995	✓	✓
40	HeNI116	Zambia	Lutembwe	F	Teeth	NHM Bulawayo	MAM 19016	1962	✓	✓
40	HeNI119	Zambia	Lutembwe	M	Teeth	NHM Bulawayo	MAM 19013	1962	✓	✓
40	HeNI134	Zambia	Lutembwe	F	Teeth	NHM Bulawayo	MAM 19011	1962	✓	✓
40	HeNI156	Zambia	Lutembwe	F	Teeth	NHM Bulawayo	MAM 19010	1962	✓	✓
41	HeNI114	Zambia	Chipangali	F	Teeth	NHM Bulawayo	MAM 19023	1962	✓	✓
41	HeNI122	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 23049	1963	✓	
41	HeNI123	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 22089	1963	✓	✓
41	HeNI128	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 23053	1964	✓	
41	HeNI132	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 22081	1962	✓	✓
41	HeNI136	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 22080	1962	✓	✓
41	HeNI141	Zambia	Chipangali	F	Teeth	NHM Bulawayo	MAM 22083	1962	✓	
41	HeNI148	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 22084	1962	✓	✓

41	HeNI150	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 23050	1963	✓	✓
41	HeNI152	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 23048	1963	✓	✓
41	HeNI153	Zambia	Chipangali	M	Teeth	NHM Bulawayo	MAM 19022	1962	✓	✓
42	HeNI133	Zambia	Katete Dam	M	Teeth	NHM Bulawayo	MAM 22088	1963	✓	✓
43	HeNI118	Zambia	Fort Jameson (Chipata)	M	Teeth	NHM Bulawayo	MAM 23452	1963	✓	✓
43	HeNI120	Zambia	Fort Jameson (Chipata)	M	Teeth	NHM Bulawayo	MAM 19024	1962	✓	✓
43	HeNI135	Zambia	Fort Jameson (Chipata)	M	Teeth	NHM Bulawayo	MAM 17452	1961	✓	✓
43	HeNI142	Zambia	Fort Jameson (Chipata)	M	Teeth	NHM Bulawayo	MAM 19015	1962	✓	✓
43	HeNI145	Zambia	Matizi (Chipata)	F	Teeth	NHM Bulawayo	MAM 17165	1962	✓	✓
43	HeNI154	Zambia	Fort Jameson (Chipata)	F	Teeth	NHM Bulawayo	MAM 19008	1962	✓	
44	He58	Zambia	Luano	—	Muscle	H. R. Siegismund	4584	—	✓	
45	HeNI124	Zambia	Zambia	M	Teeth	NHM Bulawayo	MAM 22085	1963	✓	✓
46	HeNI139	Zambia	Kasengwa (Chongwe)	F	Teeth	NHM Bulawayo	MAM 19009	1962	✓	
47	He40	Zambia	Luangwa	—	Muscle	H. R. Siegismund	2533	1995	✓	
48	HeNI191	Mozambique	Mozambique	F	Bone	NHM Berlin	ZMB_Mam_068030	—	✓	✓
49	HeNI201	Mozambique	Chifumbazi Mount.	—	Bone	NHM Berlin	ZMB_Mam_068072	—	✓	✓
50	HeNI131	Zambia	Livingstone	F	Teeth	NHM Bulawayo	MAM 5208	—	✓	✓
51	HeNI127	Zimbabwe	Gwaai	M	Teeth	NHM Bulawayo	MAM 50618	1973	✓	✓
52	HeNI115	Zimbabwe	Matobo Nat. Park	F	Teeth	NHM Bulawayo	MAM 61013	1973	✓	✓
53	HeNI130	Zimbabwe	Towla	F	Teeth	NHM Bulawayo	MAM 24428	1963	✓	✓
54	HeNI252	South-Africa	Witfonteinrant Mount.	M	Bone	Swedish NHM	NRM 581398	1846	✓	✓
54	HeNI253	South-Africa	Witfonteinrant Mount.	F	Bone	Swedish NHM	NRM 601394	1845	✓	✓
54	HeNI255	South-Africa	Witfonteinrant Mount.	F	Bone	Swedish NHM	NRM 603057	1846	✓	✓
55	HeNI232	South-Africa	South-Africa	F	Skin	NHM Wien	NMW B 5444	1842	✓	✓

56	HeNI200	South-Africa	Mashishing (Lydenburg)	—	Bone	NHM Berlin	ZMB_Mam_068070	—	✓	✓
57	HeNI137	Botswana	Sedudu, Chobe Nat. Park	M	Teeth	NHM Bulawayo	MAM 61014	1965	✓	✓
58	He31	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1358	1994	✓	
58	He32	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1359	1994	✓	
58	He33	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1360	1994	✓	
58	He34	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1361	1994	✓	
58	He35	Namibia	Mahangu Game Res.	M	Muscle	H. R. Siegismund	1362	1994	✓	
58	He36	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1363	1994	✓	✓
58	He37	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1364	1994	✓	
58	He38	Namibia	Mahangu Game Res.	F	Muscle	H. R. Siegismund	1365	1994	✓	✓d)
59	HeNI205	Namibia	Rundu	M	Skin	P-C Museum	SWA.47	1937	✓	✓
59	HeNI206	Namibia	Rundu	F	Skin	P-C Museum	SWA.48	1937	✓	✓
60	He1	Namibia	Namibia	F	Muscle	B. J. van Vuuren	Sun29	—	✓	
60	He2	Namibia	Namibia	F	Muscle	B. J. van Vuuren	Sun30	—	✓	✓
60	HeNI186	Namibia	Namibia	M	Bone	NHM Berlin	ZMB_Mam_068012	1902	✓	✓
61	HeNI183	Namibia	Grootfontein	—	Bone	NHM Berlin	ZMB_Mam_068001	—	✓	✓
62	He93	Namibia	Ovita Wildlife Restcamp	M	Muscle	H. R. Siegismund	7961	1996	✓	✓
63	HeNI184	Namibia	Kaoko	F	Bone	NHM Berlin	ZMB_Mam_068003	—	✓	✓
64	HeNI207	Angola	Cubango river	F	Skin	P-C Museum	SWA.74	1937	✓	✓
65	HeNI204	Angola	Cubango (Chitanda River)	—	Bone	NHM Berlin	ZMB_Mam_068015	—	✓	✓
66	HeNI214	Angola	Between Capelongo and Dongo rivers	F	Skin	P-C Museum	ANGI.112	1922	✓	✓
67	HeNI209	Angola	Cunene	F	Skin	P-C Museum	ANGI.56	1921	✓	✓
67	HeNI210	Angola	Cunene	M	Skin	P-C Museum	ANGI.60	1921	✓	
67	HeNI211	Angola	Cunene	F	Skin	P-C Museum	ANGI.61	1921	✓	✓
67	HeNI212	Angola	Cunene	M	Skin	P-C Museum	ANGI.63	1921	✓	✓

67	HeNI213	Angola	Cunene	M	Skin	P-C Museum	ANGI.64	1921	✓	✓
68	He25	Angola	Luando Nat. Res.	F	Muscle	P. Vaz-Pinto	24; Hn148	—	✓	
69	He26	Angola	Cangandala Nat. Park	M	Muscle	P. Vaz-Pinto	26; Hn150	—	✓	
69	He28	Angola	Cangandala Nat. Park	—	Muscle	S. Baptista	Hn343	—	✓	
70	HeNI181	Angola	Morro da Cruz	M	Bone	NHM Berlin	ZMB_Mam_067993	—	✓	✓

Sample: code He – contemporary; HeNI – historical

Country: CAR – Central African Republic

Location: Nat. Park – National Park; Res. – Reserve; Mount. – Mountain; Nat. Res. – National Reserve

Gender: M – male; F – female;

Donated by: P-C Museum – Powell-Cotton Museum, Kent, UK; NHM Wien – Natural History Museum Wien, Austria; NHM Berlin – Natural History Museum Berlin, Germany; NHM Bulawayo – Natural History Museum Bulawayo, Zimbabwe; Swedish NHM – Swedish Natural History Museum, Stockholm, Sweden

mtDNA dataset: a) through d), samples used to represent each mtDNA lineage for divergence estimations. a) north-west; b) north-east; c) east-central; and d) southern

Table S3. 2 List of 54 amplified microsatellite loci, following the recommendations in Vaz Pinto et. al. (2015) for *H. equinus*. Modifications to the mix to increase amplification success is indicated. Loci included in the dataset are indicated; if not, the reason for exclusion is provided.

Locus	GenBank accession number	Multiplex PCR			Dataset
		Original MIX	Adjusted MIX	T (º C)	
HN1	KP317498	2	2A	54	✓
HN2	KP317499	1	1A	54	✓
HN3	KP317500	4	4A	56/50	deviation from HWE
HN5	KP317501	2	2A	54	✓
HN6	KP317502	4	4A	56/50	✓
HN7	KP317503	2	2A	54	✓
HN8	KP317504	3	3A	54	✓
HN9	KP317505	1	1A	54	✓
HN10	KP317506	3	3A	54	✓
HN11	KP317507	4	4A	56/50	poor amplification
HN12	KP317508	3	3A	54	✓
HN13	KP317509	2	2A	54	✓
HN16	KP317510	3	3A	54	✓
HN17	KP317511	1	1A	54	✓
HN20	KP317512	4	4A	56/50	✓
HN21	KP317513	1	1A	54	✓
HN22	KP317514	9	5	60/55	poor amplification
HN23	KP317515	5	1	55/33	✓
HN24	KP317516	6	2	62/57	poor amplification
HN25	KP317517	5	1	55/33	✓
HN27	KP317518	8	4	60/55	✓
HN28	KP317519	8	4	60/55	✓
HN29	KP317520	8	4	60/55	✓
HN31	KP317521	5	1	55/33	✓
HN36	KP317522	8	4	60/55	✓
HN37	KP317523	9	5	60/55	✓
HN38	KP317524	5	1	55/33	✓
HN39	KP317525	5	1	55/33	poor amplification
HN41	KP317526	5	1	55/33	✓
HN45	KP317527	5	1	55/33	✓
HN46	KP317528	8	4	60/55	✓
HN47	KP317529	9	5	60/55	✓
HN48	KP317530	6	2	62/57	poor amplification
HN50	KP317531	6	2	62/57	poor amplification
HN52	KP317532	5	1	55/33	✓
HN57	KP317533	7	3	64/55	poor amplification
HN58	KP317534	9	5	60/55	✓
HN60	KP317535	6	2	62/57	poor amplification
HN61	KP317536	7	3	64/55	✓
HN64	KP317537	8	4	60/55	✓
HN68	KP317538	6	2	62/57	✓
HN72	KP317539	7	3	64/55	✓
HN75	KP317540	9	5	60/55	✓
HN79	KP317541	9	5	60/55	✓
HN80	KP317542	6	2	62/57	✓
HN81	KP317543	6	2	62/57	✓
HN86	KP317544	7	3	64/55	✓
HN91	KP317546	6	2	62/57	poor amplification
HN92	KP317547	8	4	60/55	✓
HN93	KP317548	7	3	64/55	✓
HN101	KP317549	9	5	60/55	poor amplification
HN111	KP317551	6	2	62/57	✓
HN112	KP317552	7	3	64/55	✓
HN113	KP317553	9	5	60/55	✓

Table S3. 3 Prior log-normal distributions used in Bayesian analyses for estimating mitochondrial divergence times, following Bibi (2013). Mean was set in real space, indicated in Mya (million years ago).

		Priors	
Node	**Taxa**	**Mean**	**Log(s.d.)**
1	*Alcelaphini* + *Hippotragini*	11.5	0.05
2	*Hippotragini*	6.6	0.07
3	*Alcelaphini*	6.3	0.08
4	*Alcelaphus* + *Damaliscus*	4.2	0.12
5	*Addax* + *Oryx*	3.2	0.1

Table S3. 4 The overall measure of nuclear genetic diversity for the 43 microsatellites, within-group, per locus.

	North-west		North-east		Eastern		Central		Southern	
	H_e	N_a	H_e	N_a	H_e	N_a	H_e	N_a	H_e	N_a
HN1	0.865	10	0.863	7	0.833	8	0.840	8	0.714	9
HN2	0.917	12	0.825	7	0.904	13	0.873	12	0.922	17
HN5	0.840	10	0.737	7	0.873	13	0.850	11	0.875	15
HN6	0.863	11	0.853	9	0.862	16	0.826	7	0.815	12
HN7	0.886	8	0.895	9	0.837	8	0.800	7	0.804	11
HN8	0.899	12	0.948	11	0.896	15	0.895	13	0.892	15
HN9	0.849	10	0.876	6	0.786	6	0.669	6	0.727	6
HN10	0.799	9	0.858	8	0.679	5	0.525	4	0.408	4
HN12	0.917	12	0.810	6	0.853	11	0.856	10	0.815	11
HN13	0.771	6	0.726	5	0.780	6	0.730	5	0.718	6
HN16	0.845	10	0.817	7	0.905	13	0.749	9	0.816	13
HN17	0.857	9	0.758	6	0.542	7	0.616	6	0.757	8
HN20[a)]	0.937	15	0.882	8	0.835	11	0.852	9	0.867	12
HN21	0.818	6	0.667	5	0.539	5	0.036	2	0.116	2
HN23	0.540	3	0.595	4	0.618	3	0.740	4	0.479	5
HN25	0.191	4	0.268	2	monomorphic		monomorphic		monomorphic	
HN27	0.627	3	0.542	3	0.265	3	0.578	4	0.494	4
HN28[b)]	0.827	8	0.805	5	0.723	6	0.552	4	0.766	6
HN29	monomorphic		0.189	2	0.142	2	0.032	2	0.027	2
HN31	0.053	2	0.505	2	0.506	4	0.506	2	0.491	3
HN36	0.262	2	0.268	2	monomorphic		monomorphic		0.104	2
HN37	0.832	8	0.745	5	0.769	6	0.646	6	0.811	6
HN38	0.142	2	0.521	2	0.416	2	0.275	2	monomorphic	
HN41	monomorphic		0.268	2	0.060	2	0.204	2	0.174	2
HN45	0.142	2	monomorphic		0.380	3	0.097	2	monomorphic	
HN46	0.056	2	0.451	3	0.419	3	0.316	3	0.608	4
HN47	0.534	4	monomorphic		0.108	3	0.073	2	0.155	2
HN52	0.481	2	monomorphic		monomorphic		monomorphic		monomorphic	
HN58[c)]	0.705	8	0.843	7	0.787	7	0.696	5	0.712	6
HN61	0.619	4	0.574	3	0.363	2	0.427	3	0.441	2
HN64	0.650	4	0.189	2	0.088	2	monomorphic		0.275	2
HN68	0.925	13	0.542	4	0.767	11	0.791	8	0.871	10
HN72	0.585	6	0.706	5	0.744	7	0.785	6	0.753	7
HN75	0.558	5	0.525	4	monomorphic		0.508	2	0.429	2
HN79	0.749	5	0.752	4	0.584	9	0.534	6	0.806	10
HN80	0.053	2	monomorphic		monomorphic		monomorphic		monomorphic	
HN81	0.835	9	0.850	7	0.782	7	0.837	8	0.806	10
HN86	0.894	10	0.712	5	0.638	6	0.729	7	0.435	8
HN92	0.886	9	0.874	9	0.867	9	0.840	10	0.890	14
HN93	0.852	10	0.847	8	0.857	9	0.836	11	0.836	12
HN111	monomorphic		0.100	2	0.032	2	0.033	2	0.104	2
HN112	monomorphic		monomorphic		monomorphic		monomorphic		monomorphic	
HN113	0.759	8	0.758	6	0.657	8	0.485	7	0.703	7

H_e, expected heterozygosity; **N_a**, number of alleles. All parameters were estimated allowing 5% of missing data per locus. [a)] and [c)] locus at non-HWE in Northwest population, [b)] locus at non-HWE in Central population. All significant *p-values* after Bonferroni corrections, for a $p \leq 0.01$.

Table S3. 5 Measures of nuclear genetic diversity within each roan group from the analysis of 43 microsatellites, comparing contemporary (top) and historic (bottom) samples.

CONTEMPORARY

	N	H_e ± s.d.	N_a ± s.d.	N_{pa}	AR	F_{IS}
North-west	9	0.536 ± 0.334	4.0 ± 2.3	47	3.1	0.008
North-east	6	0.555 ± 0.315	3.5 ± 1.9	12	3.1	0.015
Eastern	30	0.529 ± 0.333	5.9 ± 4.1	50	3.2	0.038
Central	3	—	—	—	—	—
Southern	13	0.497 ± 0.331	4.7 ± 3.1	34	3.0	-0.003

HISTORIC

	N	H_e ± s.d.	N_a ± s.d.	N_{pa}	AR	F_{IS}
North-west	2	0.626 ± 0.319	5.7 ± 3.2	67	4.1	0.115
North-east	4	0.551 ± 0.327	3.2 ± 1.7	12	3.9	0.091
Eastern	4	—	—	—	—	—
Central	28	0.496 ± 0.329	4.9 ± 3.3	26	3.0	0.067
Southern	23	0.533 ± 0.335	5.9 ± 4.4	46	3.8	0.165***

N, sample size; $\mathbf{H_e}$, expected heterozygosity and respective standard deviations (s.d); $\mathbf{N_a}$, average number of alleles per locus and respective standard deviations (s.d); $\mathbf{N_{pa}}$, number of private alleles; **AR**, allelic richness; $\mathbf{F_{IS}}$, inbreeding coefficient with *** $p \leq 0.001$, after Bonferroni corrections. All parameters were estimated allowing 5% of missing data per locus.

Table S3. 6 Nuclear genetic differentiation between roan population groups measured by pairwise F_{st} values (below diagonal) and standardized G'_{st} (above diagonal) from the analysis of 43 microsatellites.

	North-west	North-east	Eastern	Central	Southern
North-west	—	0.303	0.360	0.339	0.355
North-east	0.138	—	0.155	0.218	0.264
Eastern	0.163	0.070	—	0.136	0.134
Central	0.154	0.098	0.061	—	0.200
Southern	0.161	0.119	0.060	0.090	—

All values are significant with $p \leq 0.001$ based on 1,000 permutations.

Table S3. 7 Hierarchical analyses of molecular variance (AMOVA) examining the partitioning of molecular variation within and between groups of roan antelopes, for nuclear (top) and mitochondrial (bottom) markers.

NUCLEAR

	d.f.	s.s.	Variance	% Total	*p*
Among groups (N = 5)	4	152.03	0.636	10.5	< 0.000001
Among individuals within groups	126	739.43	0.423	6.9	< 0.000001
Within individuals	131	658.00	5.023	82.6	< 0.000001

MITOCHONDRIAL

	d.f.	s.s.	Variance	% Total	p
Among groups (N = 5)	4	961.07	13.18	27.0	< 0.000001
Among individuals within groups	77	2741.10	35.60	73.0	< 0.000001

d.f., degrees of freedom; **s.s.**, sum of squares; **p** values are based on 1,000 permutations.

Table S3. 8 Genetic diversity summary statistics within the entire dataset for the whole mitochondrial genome, and individually for rRNA genes and protein-coding genes.

	S	h	H_d ± s.d.	π ± s.d.
12S (954 bp)	18	16	0.725 ± 0.041	0.0014 ± 0.0010
16S (1,571 bp)	39	33	0.894 ± 0.023	0.0027 ± 0.0015
ATP6 (681 bp)	40	33	0.949 ± 0.010	0.0062 ± 0.0034
ATP8 (201 bp)	20	16	0.588 ± 0.063	0.0063 ± 0.0045
COX1 (1,545 bp)	69	45	0.971 ± 0.008	0.0050 ± 0.0026
COX2 (684 bp)	36	22	0.838 ± 0.030	0.0059 ± 0.0033
COX3 (783 bp)	36	26	0.878 ± 0.020	0.0048 ± 0.0027
CYTB (1,140 bp)	75	41	0.946 ± 0.014	0.0084 ± 0.0043
NAD1 (956 bp)	59	36	0.943 ± 0.013	0.0098 ± 0.0051
NAD2 (1,041 bp)	51	37	0.949 ± 0.012	0.0051 ± 0.0028
NAD3 (345 bp)	7	8	0.387 ± 0.068	0.0015 ± 0.0014
NAD4 (1,377 bp)	87	44	0.950 ± 0.014	0.0073 ± 0.0037
NAD4L (297 bp)	20	23	0.917 ± 0.016	0.0109 ± 0.0063
NAD5 (1,821 bp)	118	46	0.957 ± 0.014	0.0086 ± 0.0043
NAD6 (528 bp)	25	23	0.859 ± 0.027	0.0069 ± 0.0039
Total (15,551 bp)	760	70	0.993 ± 0.004	0.0059 ± 0.0028

S, number of polymorphic (segregating) sites; **h**, number of haplotypes; $\mathbf{H_d}$, haplotype diversity and respective standard deviation (s.d.); **π**, nucleotide diversity and respective standard deviation (s.d.).

Table S3. 9 Measures of genetic diversity within each roan group, based on whole mitochondrial DNA sequences, comparing contemporary (top) and historic (bottom) samples.

CONTEMPORARY

	N	S	h	H_d ± s.d.	π ± s.d.	MPD ± s.d.
North-west	3	175	3	1.00 ± 0.272	0.00750 ± 0.00562	116.670 ± 70.079
North-east	2	205	2	1.00 ± 0.500	0.01318 ± 0.13215	205.000 ± 145.310
Eastern	10	243	10	1.00 ± 0.045	0.00449 ± 0.00239	69.867 ± 32.971
Central	1	—	1	—	—	—
Southern	4	114	4	1.00 ± 0.177	0.00406 ± 0.00268	63.167 ± 34.877

HISTORIC

	N	S	h	H_d ± s.d.	π ± s.d.	MPD ± s.d.
North-west	10	418	10	1.000 ± 0.045	0.00882 ± 0.00468	131.178 ± 64.411
North-east	4	226	4	1.000 ± 0.177	0.00756 ± 0.00496	117.500 ± 64.573
Eastern	3	49	3	1.000 ± 0.272	0.00210 ± 0.00159	32.670 ± 19.877
Central	22	199	13	0.913 ± 0.045	0.00274 ± 0.00138	42.670 ± 19.249
Southern	23	221	21	0.992 ± 0.015	0.00372 ± 0.00186	57.775 ± 25.904

N, number of samples; **S**, number of polymorphic (segregating) sites; **h**, number of haplotypes; H_d, haplotype diversity and respective standard deviation (s.d.); **π**, nucleotide diversity and respective standard deviation (s.d.); **MPD**, mean pairwise distance and respective standard deviation (s.d.).

Table S3. 10 Neutrality tests' results across each major lineage, based on whole mtDNA sequences.

	Tajimas' D (p value)	Fu's F_s (p value)
North-west	-0.971 (> 0.10)	1.285 (> 0.10)
North-east	-0.657 (> 0.10)	-0.609 (> 0.10)
East - Central	-1.031 (> 0.10)	-1.859 (> 0.10)
Southern	-0.691 (> 0.10)	-2.944 (> 0.05)

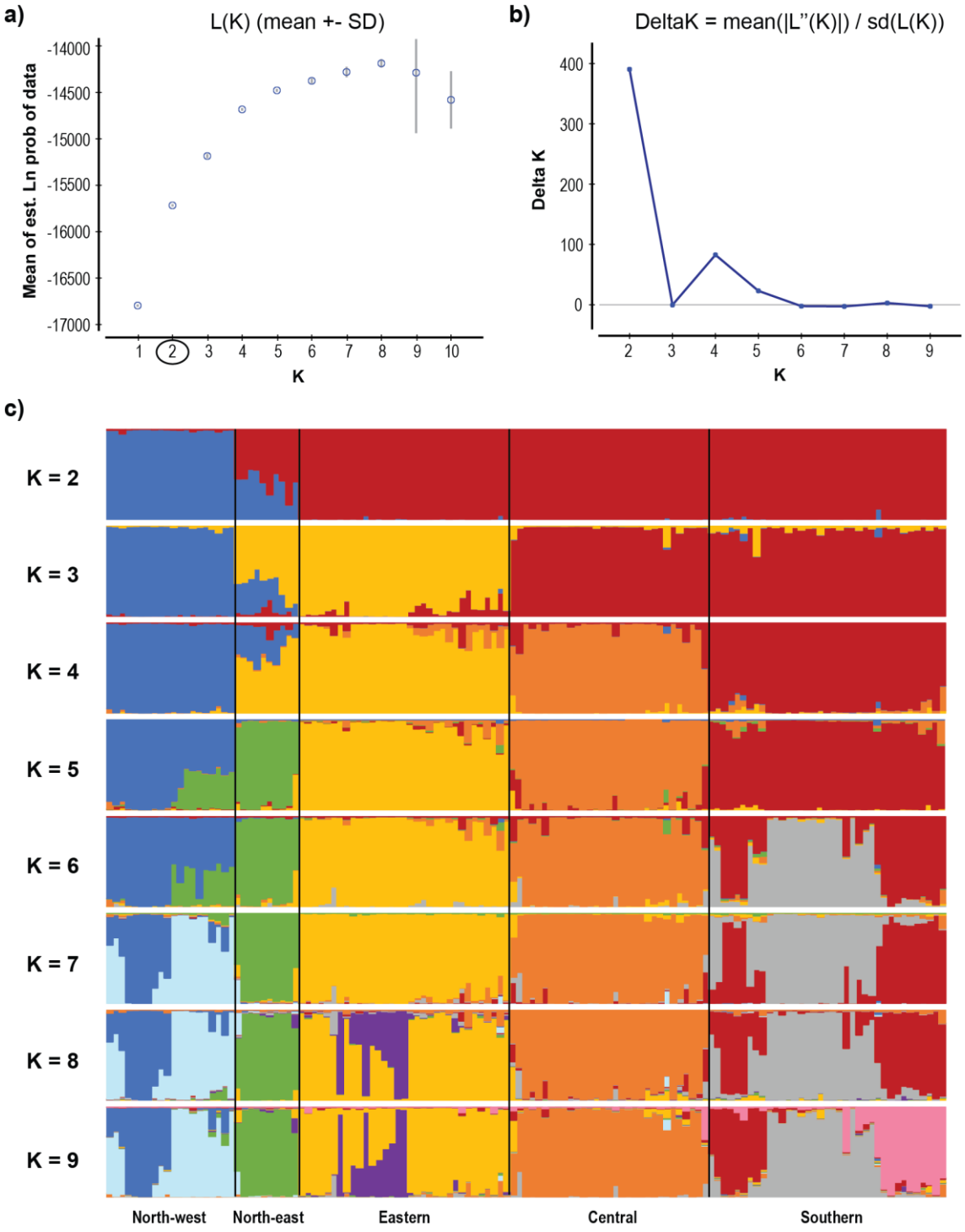


Figure S3. 1 Bayesian clustering analysis for the nuclear microsatellite data, using the complete dataset, run1. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10. Encircled K value chosen to perform subsequent hierarchical clustering analysis; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 9. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered according to sample-site number (see Table S3.1 for details) and black lines divide according to groups defined in Figure 3.1.

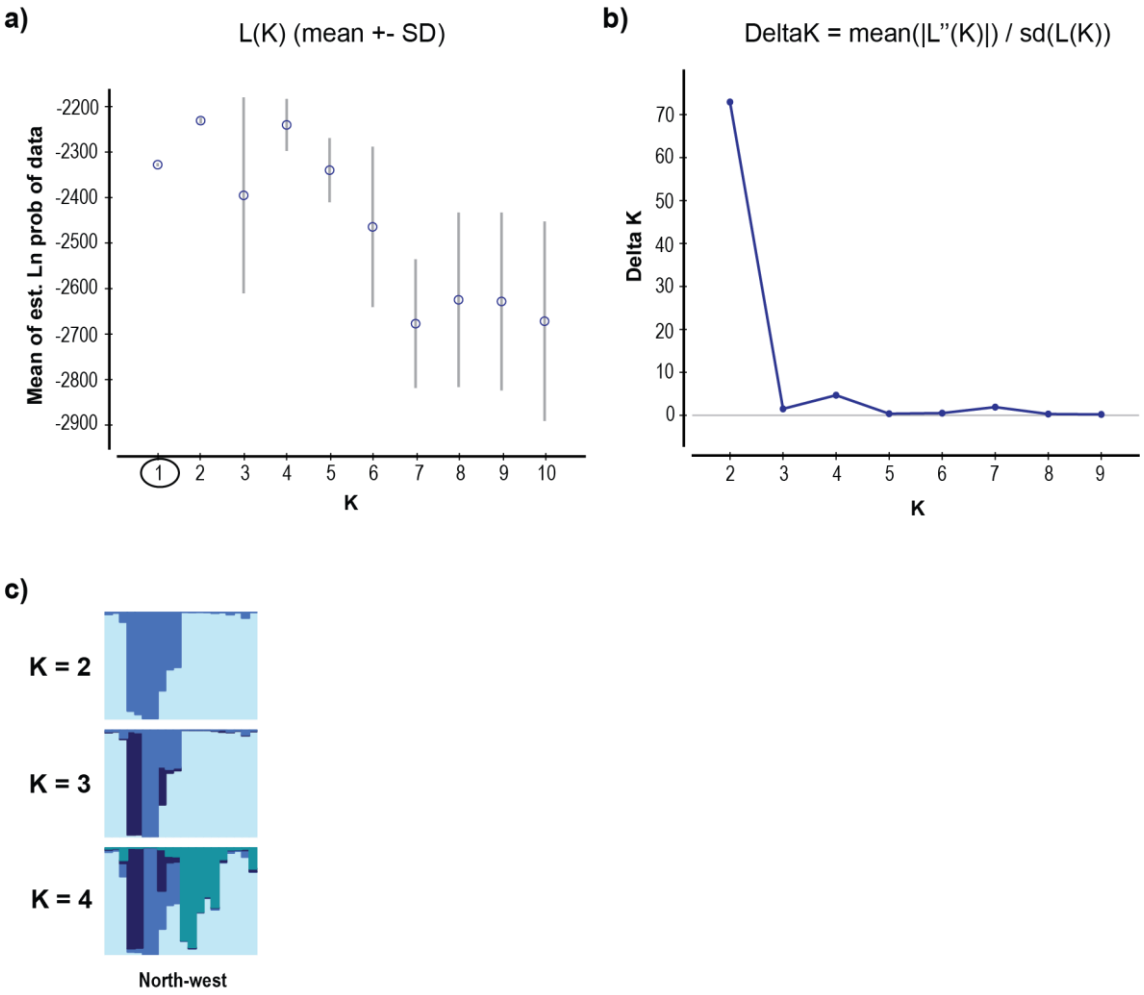


Figure S3. 2 Bayesian hierarchical clustering analysis for the nuclear microsatellite data, using cluster A, run2. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10. Encircled K value chosen as the one with the highest likelihood, based on the lowest value of K in which the L(K) curvature starts to plateau; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 4. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered according to sample-site number (see Table S3.1 for details) for the north-west group defined in Figure 3.1.

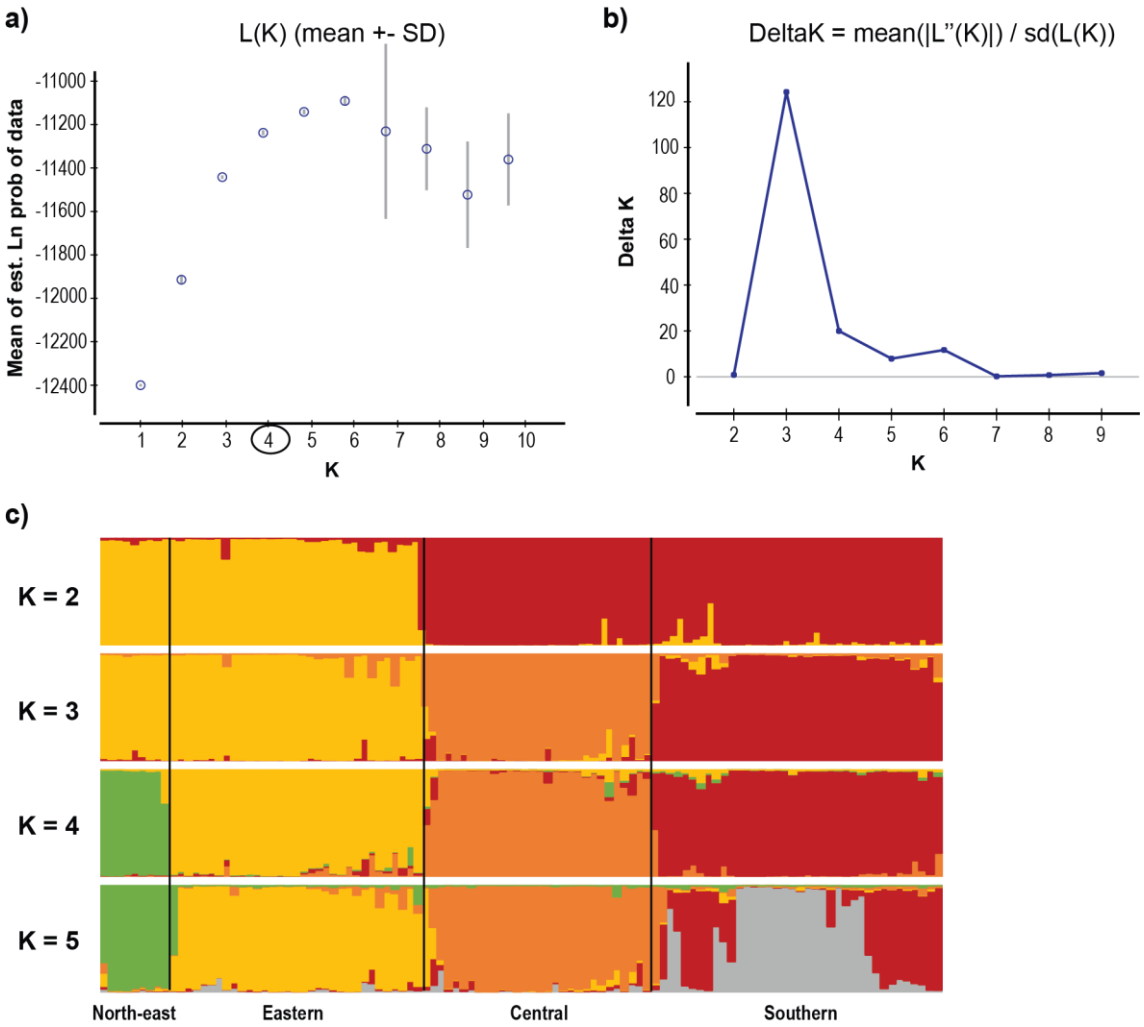


Figure S3. 3 Bayesian hierarchical clustering analysis for the nuclear microsatellite data, using cluster B, run2. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10. Encircled K value chosen as the one with the highest likelihood, based on the lowest value of K in which the L(K) curvature starts to plateau; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 5. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered according to sample-site number (see Table S3.1 for details) and black lines divide according to groups defined in Figure 3.1.

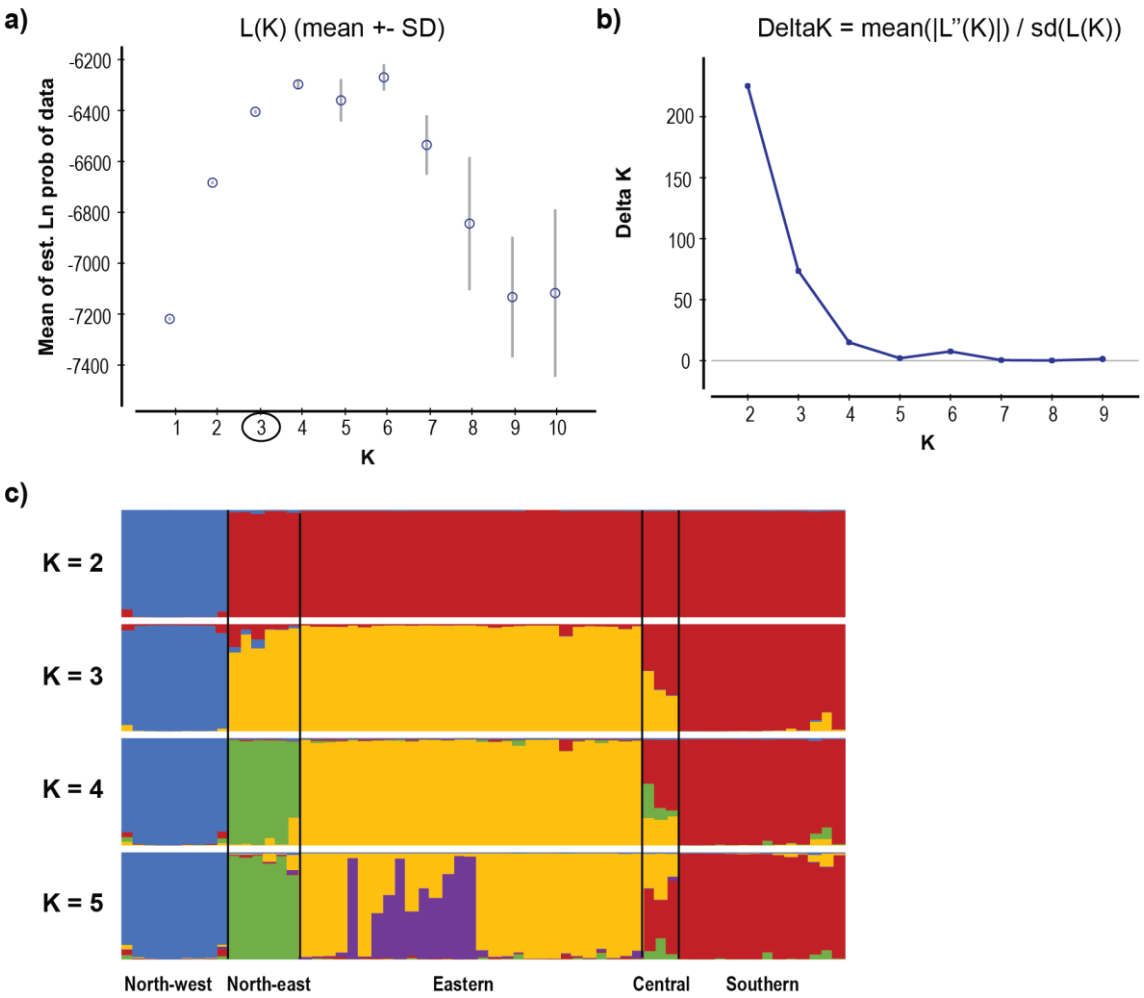


Figure S3. 4 Bayesian clustering analysis for the nuclear microsatellite data, using contemporary samples only, run3. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10. Encircled K value chosen as the one with the highest likelihood, based on the lowest value of K in which the L(K) curvature starts to plateau; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 5. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered according to sample-site number (see Table S3.1 for details) and black lines divide according to groups defined in Figure 3.1.

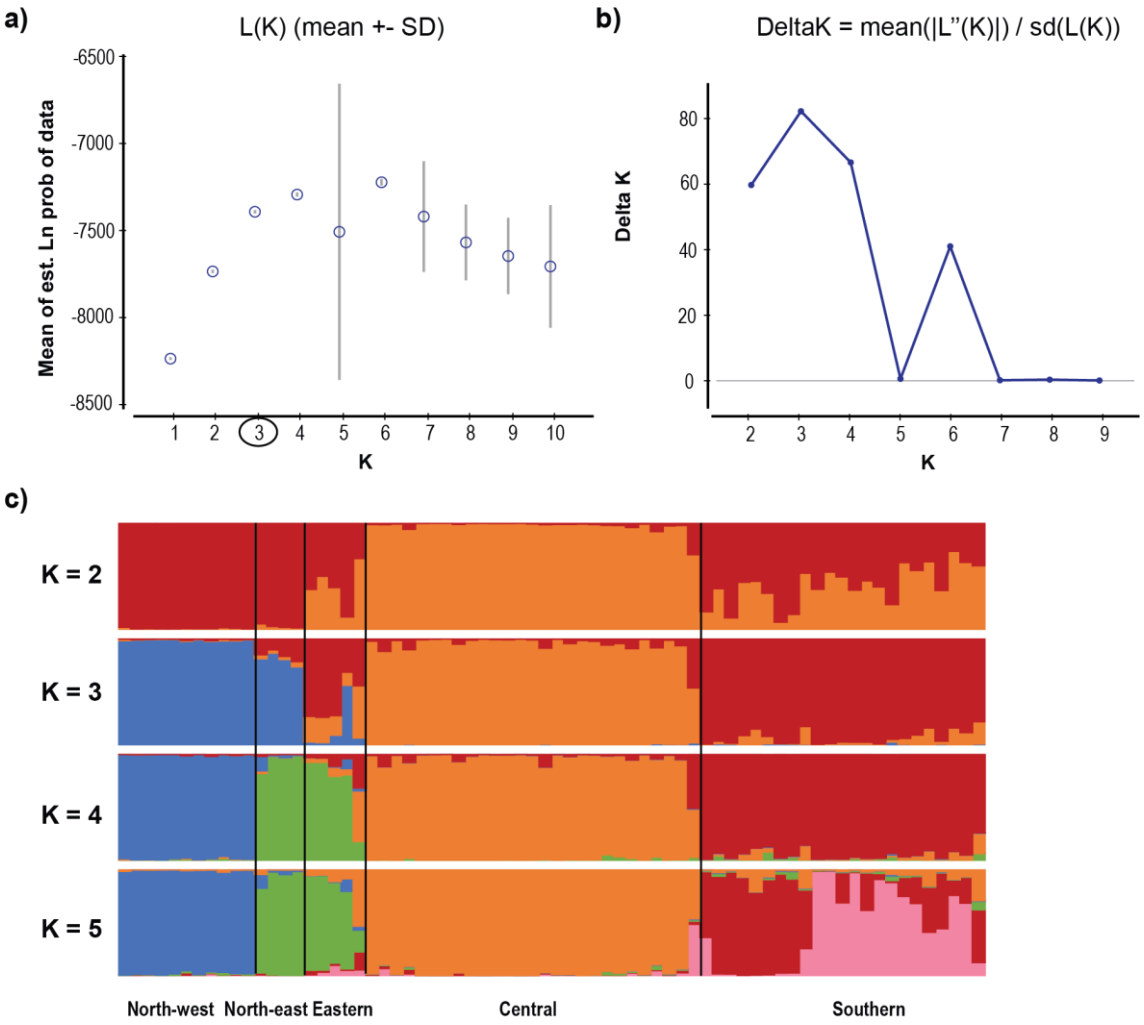


Figure S3. 5 Bayesian clustering analysis for the nuclear microsatellite data, using historical samples only, run3. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10. Encircled K value chosen as the one with the highest likelihood, based on the lowest value of K in which the L(K) curvature starts to plateau; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 5. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered according to sample-site number (see Table S3.1 for details) and black lines divide according to groups defined in Figure 3.1.

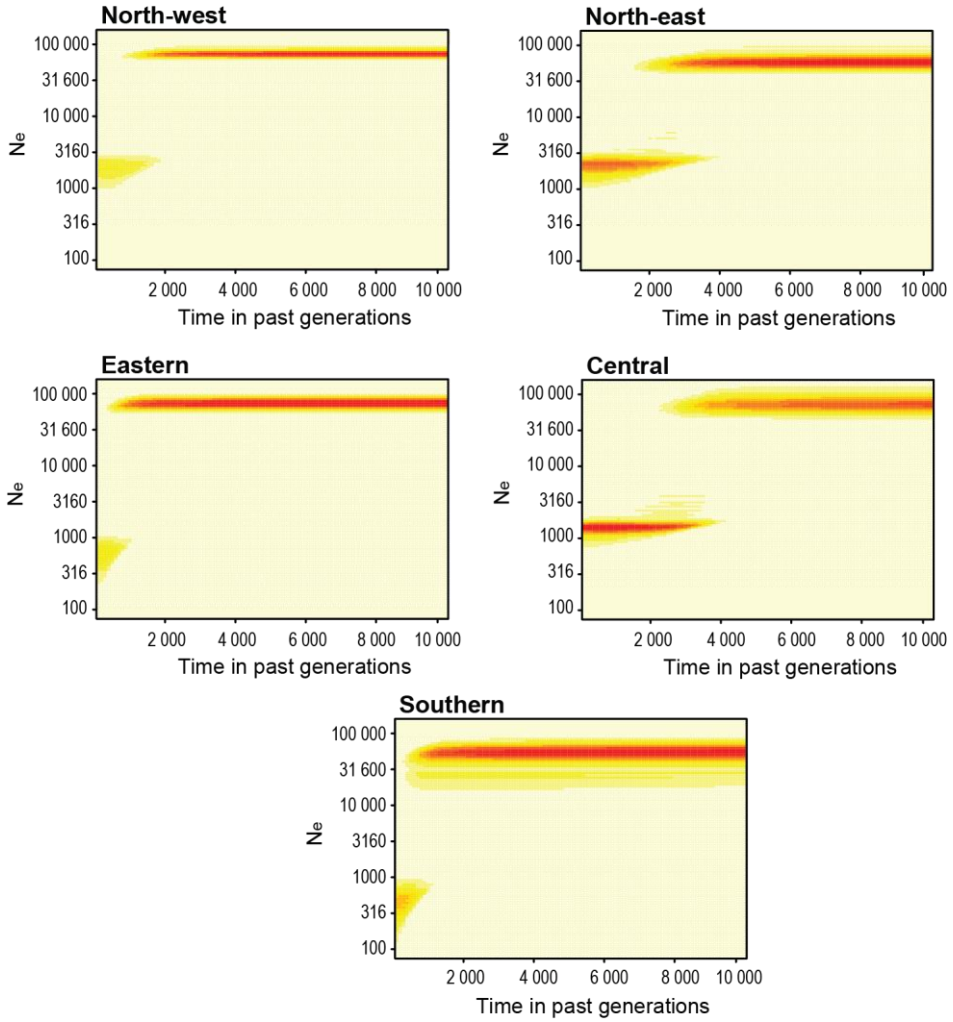


Figure S3. 6 Graphical two-dimensional summary of the posterior distribution of the estimated effective population size (N_e) for each nuclear group, using a mutation rate of $1x10^{-4}$ and two-step mutation model. The y-axis shows the N_e in logarithmic scale and x-axis shows the timescale past generations, from present (generation 0) to 10,000 generation-time. Red colour represents maximum posterior values; yellow the minimum posterior values.

Study II

De novo whole-genome assembly and resequencing resources for the roan (*Hippotragus equinus*), an iconic African antelope

Margarida Gonçalves[1,2], Hans R. Siegismund[3], Bettine Jansen van Vuuren[4], Klaus-Peter Koepfli[5,6], Nuno Ferrand[1,2,4], and Raquel Godinho[1,2,4]

Authors' affiliations: [1]CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, Portugal; [2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto; [3]Department of Biology, University of Copenhagen, Denmark; [4]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South-Africa; [5]Smithsonian-Mason School of Conservation, Front Royal, Virginia, USA; [6]Smithsonian Conservation Biology Institute, Center for Species Survival, National Zoological Park, Washington DC, USA

3.7 Introduction

Roan antelopes (*Hippotragus equinus*) are endemic to Africa, ranging throughout sub-Saharan habitats of mesic savanna woodlands. The species comprises six recognized subspecies, namely *H. e. koba* in the northwest, *H. e. charicus* and *H. e. bakeri* across central Africa, *H. e. langheldi* in the east, and *H. e. cottoni* and *H. e. equinus* in southern Africa. These subspecies were accepted by Ansell (1972) based on geographical range, but their validity has been challenged by both nuclear and mitochondrial data (Matthee and Robinson 1999; Alpers et al. 2004). To fully resolve the validity of the subspecies and regional genetic differentiation, more powerful genomic tools and sampling coverage is needed. Despite being listed as "Least Concern" by the IUCN Red List of Threatened Species, the roan antelope is severely threatened by different anthropogenic pressures, including habitat loss and epizootic diseases (IUCN SSC Antelope Specialist Group 2017). Currently, most wild populations are found only within protected areas, with decreasing population trends recorded over the past decades, some resulting in local extinctions across eastern and southern Africa (East and IUCN SSC Antelope Specialist Group 1999; Chardonnet and Crosmary 2013). Over parts of its range, the roan antelope has become established as an economically-important game species, leading to intensive captive breeding of animals. To this end, appropriately planned and executed translocations of wild animals are becoming an important management tool (Jansen van

Vuuren, Rushworth and Montgelard 2017), while the availability of reference genomes and high-resolution genomic resources add significantly to the management and conservation toolbox (Allendorf, Hohenlohe and Luikart 2010; Shafer et al. 2015; Dresser, Ogle and Fitzpatrick 2017; Brandies et al. 2019).

As part of the horse-like antelopes, the roan antelope is a member of the subfamily Hippotraginae, a group of bovids that consists of three genera namely *Hippotragus*, which includes the roan and sable antelope (*H. niger*) as well as the extinct bluebuck (*H. leucophaeus*); *Addax*, a single-species genus including only the addax (*Addax nasomaculatus*); and *Oryx*, which comprises four species, namely the beisa (*O. beisa*), the scimitar-horned oryx (*O. dammah*), the gemsbok (*O. gazelle*), and the Arabian oryx (*O. leucoryx*) (Fernández and Vrba 2005; Bibi 2013). Recently, assembled genomes became available from three species within the Hippotraginae. Studies on the gemsbok (Farré et al. 2019), the sable antelope (Koepfli et al. 2019) and the scimitar-horned oryx have exemplified how management and conservation actions can be mended by the use of genome-wide resources. Additionally, assembled reference genomes and inherent annotation information may also be used to address important biological questions related to adaptation to ecological conditions, both in natural and human-managed environments (Ge et al. 2013; Kardos et al. 2016; Armstrong et al. 2018; Martchenko et al. 2018).

In this study, we used the 10X Genomics linked-reads platform to sequence and assemble the first genome for the roan antelope. We tested its quality, comparing the roan genome assembly to other assembled genomes generated within the Hippotraginae, Bovidae, and also to other ruminant species. We also generated and used re-sequencing data of wild individuals, representing five out of the six recognized subspecies across the roan antelope's native range, to assess general levels of intraspecific diversity.

3.8 Material and Methods

3.8.1 Sample collection, library preparation and sequencing

For the reference genome, a tissue sample was collected from a roan antelope bull housed in the Lisbon Zoological Garden (Portugal; ID:10954). This individual is fifth-generation captive-bred, with a genetic pool that represents an example of *ex-situ* conservation and a collaboration between several European Zoological Gardens. The animal descends from wild-caught individuals in Uganda and Zambia, which became the founders of the population in the Hannover Zoological Garden, Germany (Figure S3.7).

The sample was collected with a tele-biopsy dart (Vario Syringes S300V, Telinject USA, Inc.), and snap-frozen in liquid nitrogen to guarantee DNA quality. The tissue sample was processed for DNA isolation, DNA library preparation and sequencing by the Genomic Services Laboratory at the HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA). Briefly, high molecular weight DNA was isolated using a Qiagen MagAttract Kit (Qiagen, Germany). Quality was determined via pulse-field electrophoresis on a Pippin Pulse system (Sage Science, USA) using the 5-Kbp–430-Kbp protocol, and concentration determined via Qubit Fluorometric Quantitation (Thermo Fisher Scientific, USA). Approximately 1 ng of DNA was used as input for Chromium Genome library preparation (v2 chemistry), which was added onto the 10X Chromium Controller (10X Genomics Inc., USA) to create Gel Bead in-Emulsions (GEMs) from natural DNA fragments. After the run, amplified fragments within each GEM were pooled into anchored longer fragments, according to molecular barcodes, and used for subsequent steps (van Dijk et al. 2018). Post libraries were constructed using standard Chromium Genome i7 indexes, quantified by qPCR (Kapa Biosystems, USA) and sequenced on 2 lanes of an Illumina HiSeq X Ten system flow cell (Illumina, USA), generating 150-bp paired-end reads.

For whole genome re-sequencing, we used five tissue samples from wild individuals, representing five out of the six currently recognized subspecies for the roan antelope (Figure 3.5; Table 3.3). Three samples were contemporary muscle preserved in ethanol 96% from the tissue collection of the University of Copenhagen, whereas the remaining two samples were historic dried skin donated by the Powell-Cotton Museum (Kent, UK). Total genomic DNA for contemporary samples was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Germany) following the manufacturer's instructions. For the historic samples, after re-hydration, DNA extraction was performed in a laboratory dedicated to low-quality DNA following the protocol of Dabney et al. (2013). Negative controls were used throughout both DNA extraction processes to monitor for potential contamination. DNA of contemporary samples was sheared by sonication and fragments of ca. 350 bp were selected using AMPure XP bead clean-up protocol (Beckman Coulter, USA). For historic samples, no fragment size-selection was performed. Historic DNA was incubated with USER (Uracil-Specific Excision Reagent) enzyme (NEB, New England Biolabs, UK) for uracil excision. Double-stranded DNA library preparation followed Meyer and Kircher (2010) and Kircher et al. (2012) protocols for dual-indexing. Amplification and purification were done following Dabney and Meyer (2012), after determining by qPCR the number of amplification cycles that minimizes hydrolytic damage (Swillens et al. 2004; Stiller et al. 2006) (Table 3.3). Amplified libraries from both contemporary and historic samples were cleaned using MinElute PCR purification kit (Qiagen, Germany),

quantified by Kapa Library Quantification Kit (Roche Sequencing and Life Science, USA), and pooled at equimolar amounts. Pooled sequencing was performed on a HiSeq 4000 platform (Illumina, USA) using the 150-bp paired-end sequencing protocol.

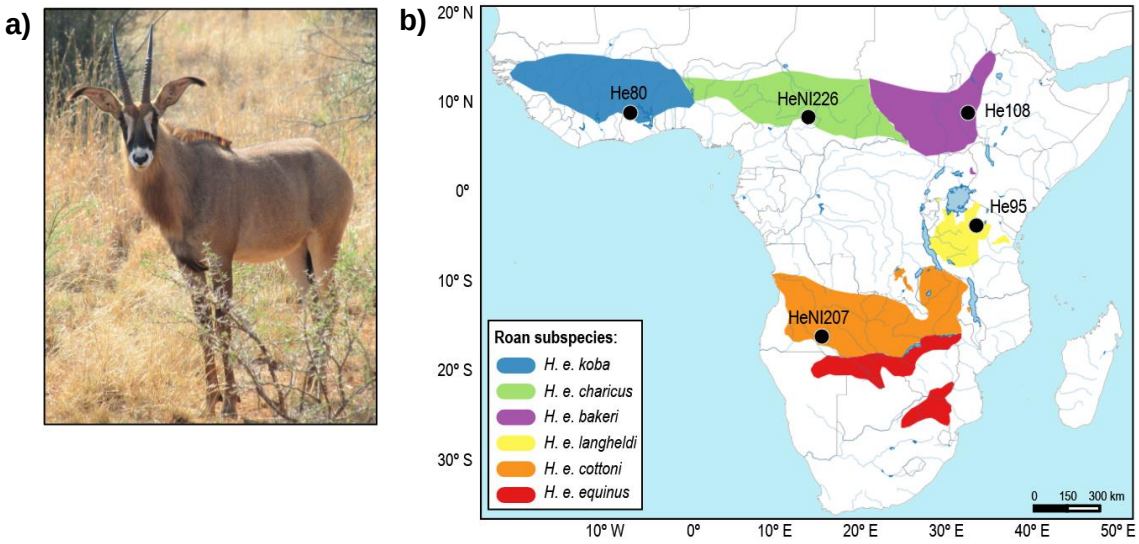


Figure 3. 5 Roan antelope distribution and sampling. a) Roan antelope (*Hippotragus equinus*). Photo by Raquel Godinho; b) Geographic distribution of roan antelope subspecies according to Ansell (1972) and current species distribution, following IUCN SSC Antelope Specialist Group (2017). The geographic location of the five samples used in this study for whole genome re-sequencing is represented by dark dots labelled with sample identifications (as in Table 3.3).

Table 3. 3 Roan antelope native samples used for whole genome re-sequencing.

Sample ID	Original label ID	Subspecies	Year	Gender	Locality	Nº PCR cycles	Mapping (%)	Mean coverage
He80	7324[a]	*H. e. koba*	1998	F	Kablima, Ghana	8	99.8	5.2x
HeNI226	NN.252[b]	*H. e. charicus*	1925	M	Nana Barya Reserve, CAR	9	99.7	5.9x
He108	9293[a]	*H. e. bakeri*	—	M	Akobe septum, Ethiopia	8	99.8	11.7x
He95	8250[a]	*H. e. langheldi*	1997	M	Maswa Game Reserve, Tanzania	10	99.7	12.2x
HeNI207	SWA.74[b]	*H. e. cottoni*	1937	F	Cubango river, Angola	7	99.7	14.1x

[a] contemporary samples from tissue collection of the University of Copenhagen, Department of Biology (responsible: H.R. Siegismund); [b] historic samples donated by the Powell-Cotton Museum, Kent, UK; CAR – Central African Republic.

3.8.2 Genome assembly and completeness

Total raw-data was assembled using the 10X Genomics software Supernova v.2.1.1 (Weisenfeld et al. 2017). We used ca. 1.2 Gbp randomly selected reads to obtain a raw read coverage of 56x, as recommended by 10X Genomics. Subsequent fasta files were generated using Supernova mkoutput, with a kmer junction of k = 48, a minimum contig size of 1,000 bp, and using the raw style, where all edges are represented by a fasta record. Raw style was compared to two additional style outputs, where branches are selected according to coverage, gaps are merged with subsequent sequences, and no reverse complement sequences are represented: 1) megabubble style, where each arm is represented by a fasta record, and 2) pseudohap style, which extracts a single pseudohaplotype per scaffold, choosing arbitrarily between maternal and paternal alleles. Duplicated scaffolds were removed with GenomeTools sequniq v.1.6.1 (Gremme, Steinbiss and Kurtz 2013).

Assembly completeness was assessed in Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3.0.2 (Simão et al. 2015), using the Mammalia OrthoDB v9 gene set (Zdobnov et al. 2017) containing 4,104 genes. By comparing BUSCO results between different output assembly styles, we found that the pseudohaplotype fasta style had the highest completeness (Figure S3.8) Therefore, only this assembly was retained for further analyses.

3.8.3 Genome annotation and synteny

To analyse the repeat content of the roan genome, we generated a *de novo* repeat library from the assembly using RepeatModeler v.2.0.1 (Zeng et al. 2018) which integrates both RECON v.1.08 (Bao and Eddy 2002) and RepeatScout v.1.0.6 (Price et al. 2005) to predominantly find transposable elements. *De novo* generated interspersed repeats were classified using the integrated script in RepeatClassifier (Zeng et al. 2018). Based on this custom library, the roan assembly was screened for repetitive elements using RepeatMasker v.4.0.9 (Smit, Hubley and Green 2015) with the NCBI-RMBlast v.2.6.0+ search engine, including interspersed genomic repeats and low-complexity sequences. We used the -s option to increase sensitivity and -xsmall to produce a soft-masked output, with repeat regions indicated in lower case letters.

We used Augustus v.3.3.2 (Stanke et al. 2008) to identify candidate protein-coding genes in the masked-assembled genome. The software was launched disabling annotation of untranslated regions (--UTR=off), using the masked sequence as evidence against exons (--softmasking=1) and applying the human gene model for gene prediction (–species=human). Candidate genes were translated into protein sequences using the incorporated Augustus script getAnnoFasta and filtered using eggNOG-mapper (Huerta-

Cepas et al. 2017). Annotation quality was assessed based on eggNOG v.4.5 orthology data (Huerta-Cepas et al. 2016), with a minimum query cover of 50% and an e-value cut-off of $1x10^{-4}$. Finally, annotation quality of the roan genome was compared with two other Hippotraginae (sable and scimitar-horned oryx) and the domestic cow assemblies for a set of common metrics using GenomeQC (Manchanda et al. 2020).

Roan antelope genome assembly was aligned to the domestic cow genome (BosTau9 – GenBank accession number: GCA_002263795.2; Zimin et al. 2009) using LAST v.0.874 (Kiełbasa et al. 2011). The domestic cow genome was first prepared for alignment using the lastdb command option, and the alignment was run using lastal and last-split commands. Then, we used the maf-swap incorporated script to change the order of the sequences into the resulting MAF-format alignment to obtain the best pairwise synteny blocks between genomes. Ordered scaffolds above 10 Kbp were used to visualise genome synteny with Circos v.0.69.6 (Krzywinski et al. 2009).

3.8.4 Re-sequencing alignment

Read data from the five re-sequenced wild individuals were aligned to the roan genome assembly using BWA-mem (Li 2013) with default settings, following adapter-trimming with Trimmomatic v.036 (Bolger et al. 2014) and retaining only reads > 50 bp. Duplicated reads were marked with Picard v.2.21.4 (http://broadinstitute.github.io/picard/) and local realignment around indels was improved using GATK v.3.8 IndelRealigner (McKenna et al. 2010). Final mapping quality was assessed with QualiMap v.2.2.2 (Okonechnikov, Conesa and García-Alcalde 2016).

3.8.5 SNV calling and filtering

Filtered alignments were used in GATK v.3.8 HaplotypeCaller (van der Auwera et al. 2013) to separately call potential single-nucleotide variants (SNVs) for each of the five individuals. GATK GenotypeGVCFs was then used for joint genotyping based on individual genomicVCFs. Joint genotyping on cohorts improves variant calling by preventing base uncertainty errors. Using both BCFtools v.1.9 (Li et al. 2009) and VCFtools v.0.1.16 (Danecek et al. 2011), variant data was filtered to retain 1) only bi-allelic SNVs, 2) SNVs without missing data, 3) no indels, 4) SNVs covered by more than 5 reads, and 5) SNVs with a quality score > 50. The final set of filtered SNVs was used for principal component analysis, with the gdsfmt v.1.8.3 and SNPRelate v.1.6.4 R packages (Zheng et al. 2012).

3.9 Results and Discussion

3.9.1 Genome assembly and completeness

The genome sequencing of roan individual ID:10954 using 10X Genomics Chromium generated approximately $1.2x10^9$ paired-end reads (Table S3.11). The raw and effective (i.e. the number of reads retained after filtering) read coverage recommended by 10X Genomics is 56x and 42x, respectively. The assembly summary statistics confirm the obtained values as close to the recommended ones (60.7x raw and 39.7x effective coverage; Table S3.11). This genome has a contig N50 of 239.6 Kb, assembled into 16,880 scaffolds with an N50 of 8.42 Mb (Table S3.11). Overall assembly quality is reflected in the high number of scaffolds (> 2,200) more than 10 Kb in length and in the low percentage (4.1%) of missing bases from such scaffolds (Table S3.11). Total genome assembly size was 2.56 Gb, which is close to the genome size estimated from the *k-mer* distribution of 3.01 Gb (Table S3.11). The Supernova software estimated a genome-wide G+C content of 41.9%, and an overall heterozygosity of 0.0015 (one heterozygous position every 636 bp).

The assembly size and the G+C content of our roan antelope genome is comparable to other available genomes for Hippotraginae, as well as to genomes of other members of the Bovidae family and other ruminant species. For example, the assembly genome size for the European bison and the sable antelope are 2.58 Gb and 2.60 Gb, respectively (Wang et al. 2017; Koepfli et al. 2019), whereas estimates of G+C content for the sable antelope and the scimitar-horned oryx are 41.8% (Koepfli et al. 2019; Humble et al. 2020). However, the 10X Genomics Chromium library system implemented for the roan genome assembly allowed us to considerably improve the scaffold N50 (8.42 Mb) in comparison with other strategies used for the sable antelope (4.59 Mb; Koepfli et al. 2019), the blue wildebeest (3.5 Mb; Chen et al. 2019), the Tibetan antelope (2.76 Mb; Ge et al. 2013), or for the African buffalo (2.32 Mb; Chen et al. 2019). Such improvements probably reflect the unique 10X Genomics library preparation protocol, which was specifically created to obtain long-range anchored information (van Dijk et al. 2018). The mean heterozygosity of 0.0015 was in line with estimates within mammal species and is higher than the value observed for Hippotraginae of critical conservation concern (scimitar-horned oryx; Humble et al. 2020).

BUSCO evaluation of gene completeness retrieved 91.2% complete genes (3,743 out of 4,104 genes; Table 3.4). This value is similar to the one observed for the Thomson's gazelle (91.1%) by Chen et al. (2019), but lower than the values retrieved for the

gemsbok (above 92%) and the sable antelope (94.8%) by Farré et al. (2019) and Koepfli et al. (2019), respectively.

Table 3. 4 Assembly statistics based on gene completeness scores by BUSCO v.3.0.2 for the whole genome assembly of roan antelope.

BUSCOs	Total number	Percentage
Complete	3,743	91.2%
Complete and single-copy	3,677	89.6%
Complete and duplicated	66	1.6%
Fragmented	194	4.7%
Missing	167	4.1%

BUSCO scores using the Mammalia OrthoDB v9 gene set, containing 4,104 genes.

3.9.2 Genome annotation and synteny

An estimated 42.2% (about 1.1 Gbp) of the roan genome was composed of repetitive sequences, based on short- and long-interspersed nuclear elements (SINEs and LINEs), long-interspersed retrotransposons (LTR), DNA elements, small RNAs, and simple and low complexity tandem repeats (Table 3.5). LINEs were the most common repetitive element, representing 30.4% of the overall content, followed by LTR elements, which represented 4.0%. We also obtained a low percentage of unclassified repeats (2.5%) that did not correspond to any familiar element in the RepeatModeler specific library. Transposable elements exist in elemental families and comprise a significant fraction of eukaryotic genomes (Biscotti, Olmo and Heslop-Harrison 2015). Each family is derived from the multiplication of a particular element in the genome of a common ancestor and is thus a proxy for shared evolutionary history across species (Schibler et al. 2006; Platt, Vandewege and Ray 2018; Qi et al. 2018). Thus, a similar percentage of the repetitive content among assemblies of roan antelope and other members of its evolutionary lineage were expected (e.g., Ou, Chen and Jiang 2018). Accordingly, the 42.2% of repetitive sequences observed for the roan assembly is consistent with the genome assemblies of other Hippotraginae, such as the sable antelope (46.97%), the gemsbok (41.27%), and the scimitar-horned oryx (47.8%) (Farré et al. 2019; Koepfli et al. 2019; Humble et al. 2020), and also other bovids as the domestic cow (45.28%) and the European bison (47.3%) (Zimin et al. 2009; Wang et al. 2017). Additionally, long-interspersed repeats (LINEs) were the most common element found in our assembly, which is also in agreement with general knowledge for the bovine genome (Adelson, Raison and Edgar 2009). However, identification of transposable elements can be achieved through numerous available tools, and discrepancies observed across different

genomes may be related to genome assembly quality and different challenges during the identification process (Saha et al. 2008; Makałowski et al. 2019).

Table 3. 5 Summary of the repetitive content found in roan antelope genome assembly, using RepeatMasker.

	Number	Length occupied (bp)	Percent masked
SINEs	222,376	27,885,908	1.1
LINEs	2,731,480	799,087,834	30.4
LTR elements	496,668	106,731,053	4.0
DNA elements	377,402	54,829,227	2.1
Unclassified	413,329	65,638,505	2.5
Total interspersed repeats		1,054,172,527	40.1
Small RNA	298,364	38,764,415	1.4
Simple repeats	470,855	18,232,272	0.6
Low complexity	80,149	3,946,798	0.1
Total repeats		1,106,510,938	42.2

De novo genome annotation with Augustus identified a total of 30,622 candidate protein-coding genes (PCGs), of which 20,518 matched common gene names, after filtering with eggNOG-mapper. Direct comparisons of these values with other assemblies could be biased, since annotation and the final number of estimated PCGs can be influenced by the assembly method and quality (Florea et al. 2011). We avoided this possible bias using the recently available web framework GenomeQC that enables comparison across multiple assemblies and assembly types. The final number of PCGs estimated for our roan assembly is consistent with estimates for the domestic cow (30,589; Zimin et al. 2009), and slightly higher than those estimated for both the sable antelope (23,846) and the scimitar-horned oryx (28,559) genome annotations (Koepfli et al. 2019; Humble et al. 2020). Such estimates can be used as proxy for the quality of the roan genome (Table S3.12).

Roan scaffolds alignment against the domestic cow reference genome resulted in 85.6% synteny alignment (ca. 2.25 Gbp), with an average identity of 92.7%. Long stretches of synteny blocks covered more than 95% of each of the 29 bovine autosomal chromosomes, as well as the X chromosome (Figure 3.6; Table S3.13).

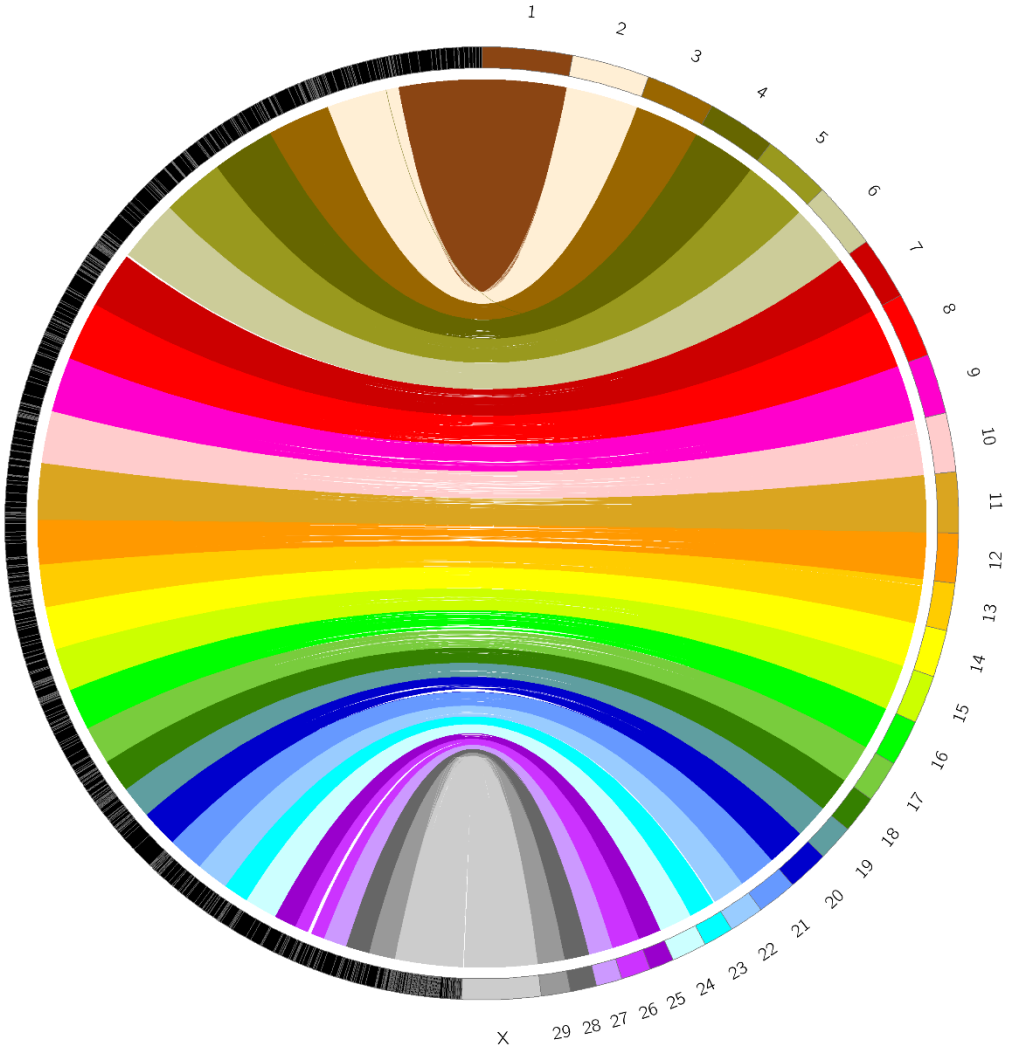


Figure 3. 6 Synteny blocks between the roan antelope and the domestic cow genomes. Synteny between roan scaffolds larger than 10 Kbp and the domestic cow chromosomes (29 autosomes and the X). Scaffolds were ordered to obtain the best pairwise alignment blocks between the two genomes which are visualized using coloured lines.

3.9.3 Intraspecific diversity

Alignment of the five re-sequenced wild individuals to the roan antelope genome assembly resulted in 99.7% and 99.8% of mapped reads (Table 3.3). The average depth of coverage was 9.8x, ranging between 5.2x and 14.1x for Ghana and Angolan individuals, respectively (Table 3.3). Our intraspecific evaluation of the roan genome allows the first assessment of genomic diversity patterns of the species across its native distribution range. From the alignment of the re-sequenced individuals, we found a total of 21,127,946 SNVs. Among these, 20,896,201 were bi-allelic which, after quality filtering, resulted in a final set of 3,401,741 SNVs across the five roan antelope individuals. Overall transition/transversion ratio across SNVs was 2.07 (2,294,647 / 1,107,094). The number of alternative homozygous SNVs varied between 949,845 and

577,765 SNVs for individuals from Ghana and Tanzania, respectively. The number of heterozygous SNVs ranged from 1,043,928 in He108 from Ethiopia to 711,962 in HeNI207 from Angola (Figure 3.7a). Results for historic samples did not show any obvious bias for low coverage, reduced alignment efficiency or low number of heterozygous SNVs, which can be an important source of errors for degraded and chemically modified DNA samples (Rowe et al. 2011; Ewart et al. 2019). However, sampling size in our study hampers further interpretations.

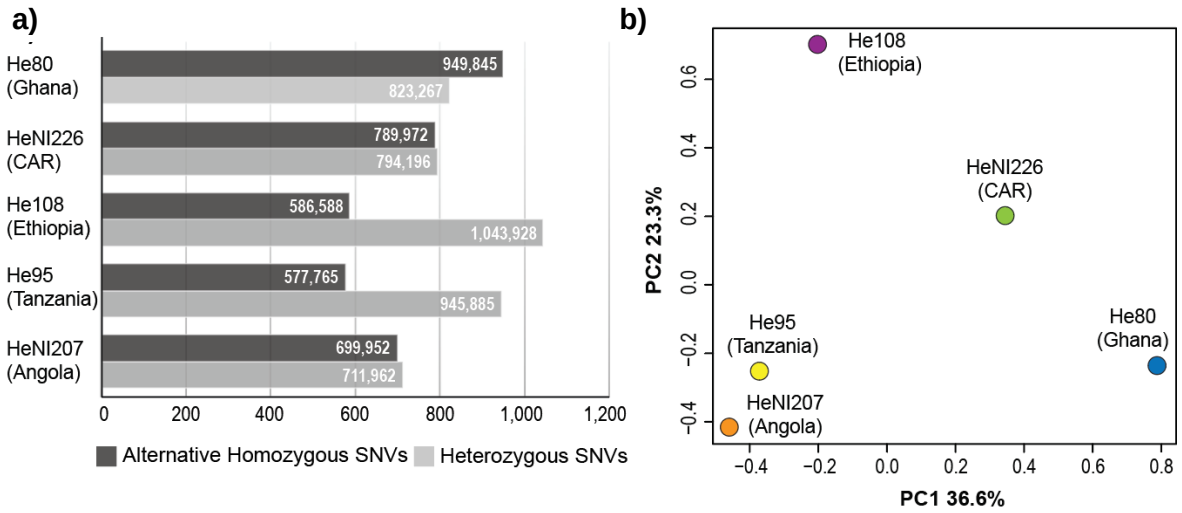


Figure 3. 7 Intraspecific variation in roan antelopes. **a)** Bar chart comparing the number of quality-filtered alternative homozygous and heterozygous single-nucleotide variants (millions of SNVs) among the five re-sequenced wild roan individuals. **b)** Plot of principal component analysis (PCA) for the same individuals. PC1 and PC2 show the first and second axes with corresponding percentage of explained variance. Circles identify each individual and colours correspond to the respective subspecies, following Figure 3.5.

Principal component analysis provided a visualization of genetic distances among re-sequenced individuals, which correlates with their geographical distribution (Figure 3.7b). The two first axes explain 60% of the total variance observed among the genome of these individuals. The first axis (PC1) explains 36.6% of the variance and separates representatives in the northern distribution of the range, namely from Ghana and Central African Republic, from those in the central and southern range of the species, namely Ethiopia, Tanzania and Angola. The second axis (PC2), with an explained variance of 23.3%, retrieved the individual dispersion along the west-east geographical axis. Therefore, levels of intraspecific differentiation translate into a clear separation between north-south and west-east geographical axes. Although based on only a few individuals, this result supports the currently recognized subspecies and previous results on roan antelope population genetics, including the two proposed Evolutionary Significant Units (ESUs) (Ansell 1972; Alpers et al. 2004). Interestingly, we found a higher number of heterozygous SNVs for Ethiopian (~$1.0x10^6$) and Tanzanian (~$9.4x10^5$) individuals

among the five wild individuals, indicative of a higher genetic diversity. Increased intraspecific genetic diversity may be, among other factors, a signature of the presence of a contact zone, following population diversification in isolation and range expansions (Hewitt 2011). Roan populations in Ethiopia and Tanzania are located in a possible contact zone between the proposed ESUs, which would extend towards west and south of those regions, respectively. Additionally, the higher proportion of reference alleles in these two individuals also indicates a closer genetic similarity to the assembled genome. This may also be explained by the evidence that the sample used to build the genome assembly is from a five-generation captive-bred individual with a maternal ancestry of wild-caught individuals from Uganda and Zambia (Figure S3.7). Both countries are geographically closer to Ethiopia and Tanzania, and therefore, higher genetic similarity to the reference genome may reflect geographical proximity of these populations. Studying genome-wide diversity constitutes an important application of SNVs towards species management in both ex- and in-situ programmes, as it is being applied for the sable antelope (Gooley et al. 2020).

3.9.4 Future prospects

Non-model mammal species are difficult to sample in the wild due to cost and/or logistical constraints (Etherington et al. 2020). *Ex-situ* management programmes, such as the ones carried out by zoological gardens, can become important donors of unique genomic resources (notably fresh tissue or non-invasive samples for DNA) for threatened and non-threatened species (Clarke 2009; Norman, Putnam and Ivy 2019). From a sample collected at the Lisbon Zoological Garden, we successfully assembled the first genome for the roan antelope, one of the most iconic dwellers of the African savanna. This draft genome assembly represents a valuable genomic resource that may provide input, among others, on phylogenetic relationships, demographic history, and evolution of adaptive traits, such as headgear, multichambered stomach, and adaptation to extreme environments (Bovine Genome Sequencing and Analysis Consortium et al. 2009; Reese et al. 2010; Bao, Lei and Wen 2019; Chen et al. 2019). Across the three genera that comprise Hippotraginae, both *Oryx* and *Addax* species are arid-adapted, whereas *Hippotragus* species dwell in more mesic savanna habitats. The adaptation to different environments is an example of a future research topic leading to important biological discoveries among this subfamily. Furthermore, the availability of the roan's reference genome and the additional genomic resources included in this work may prove highly valuable on management and conservation decisions (Dresser, Ogle and Fitzpatrick 2017; Supple and Shapiro 2018; Brandies et al. 2019). Improvements to the

roan's draft genome can be accomplished by proximity ligation sequencing (Hi-C) to generate chromosome-length scaffolds (Dudchenko et al. 2018).

3.10 References

Adelson, D. L., Raison, J. M. and Edgar, R. C., 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proceedings of the National Academy of Sciences*, 4: 12855–12860.

Allendorf, F., Hohenlohe, P. and Luikart, G., 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11: 697–709.

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Ansell, W. F. H., 1972. Order Artiodactyla, pp. 15–83 in *Mammals of Africa: an Identification Manual*. Smithsonian Institute Press (J. Meester and H. W. Setzer, Eds.), Washington, DC.

Armstrong, E. E., Taylor, R. W., Prost, S., Blinston, P., Van Der Meer, E., Madzikanda, H., Mufute, O., Mandisodza, R., et al., 2018. Entering the era of conservation genomics: cost-effective assembly of the African wild dog genome using linked long reads. *Biorxiv*, *195180*.

Bao, W., Lei, C. and Wen, W., 2019. Genomic insights into ruminant evolution: from past to future prospects. *Zoological Research*, 40: 476-487.

Bao, Z. and Eddy, S. R., 2002. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genetics Research*, 12: 1269–1276.

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

Biscotti, M. A., Olmo, E. and Heslop-Harrison, J. S., 2015. Repetitive DNA in eukaryotic genomes. *Chromosome Research*, 23: 415–420.

Bolger, A. M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30: 2114–2120.

Bovine Genome Sequencing and Analysis Consortium, Elsik, C., Tellam, R. and Worley, K., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324: 522–528.

Brandies, P., Peel, E., Hogg, C. J. and Belov, K., 2019. The value of reference genomes in the conservation of threatened species. *Genes*, 10: 846.

Chardonnet, P. and Crosmary W., 2013. *Hippotragus equinus* Roan Antelope, pp. 548–556 in *Mammals of Africa* VI. Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Bibi, F., Yang, Y., Wang, J., et al., 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, 364: eaav6202.

Clarke, A. G., 2009. The Frozen Ark Project: the role of zoos and aquariums in preserving the genetic material of threatened animals. *International Zoo Yearbook*, 43: 222–230.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, 52: 87–94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., et al., 2011. The variant call format and VCFtools. *Bioinformatics*, 27: 2156–2158.

Dresser, C., Ogle, R. and Fitzpatrick, B., 2017. Genome scale assessment of a species translocation program. *Conservation Genetics*, 18: 1191-1199.

Dudchenko, O., Shamim, M., Batra, S., Durand, N., Musial, N., Mostofa, R., Pham, M., St Hilaire, B. G., et al., 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *Biorxiv*, 254797.

East, R. and IUCN SSC Antelope Specialist Group, 1999. African antelope database 1998 (IUCN/SSC Antelope Specialist Group, Ed.), Gland: IUCN.

Etherington, G., Heavens, D., Baker, D., Lister, A., McNelly, R., Garcia, G., Clavijo, B., Macaulay, I., Haerty, W. and di Palma, F., 2020. Sequencing smart: de novo sequencing and assembly approaches for a non-model mammal. *GigaScience*, 9: giaa045.

Ewart, K. M., Johnson, R. N., Ogden, R., Joseph, L., Frankham, G. J., Lo, N., 2019. Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species. *Molecular Ecology Resources*, 19:1578–1592.

Farré, M., Li, Q., Zhou, Y., Damas, J., Chemnick, L., Kim, J., Ryder, O. A., Ma, J., et al., 2019. A near-chromosome-scale genome assembly of the gemsbok (Oryx gazella):

an iconic antelope of the Kalahari desert. *GigaScience*, 8: giy162.

Fernández, H. and Vrba, E., 2005. A complete estimate of the phylogenetic relationships in Ruminantia: A dated species-level supertree of the extant ruminants. *Biological Reviews*, 80: 269–302.

Florea, L., Souvorov, A., Kalbfleisch, T. S., Salzberg, S. L., 2011. Genome assembly has a major impact on gene content: a comparison of annotation in two *Bos taurus* assemblies. *Plos One*, 6: e21400.

Ge, R., Cai, Q., Shen, Y., San, A., Ma, L., Zhang, Y., Yi, X., Chen, Y., et al., 2013. Draft genome sequence of the Tibetan antelope. *Nature Communications*, 4: 1–7.

Gooley, R. M., Tamazian, G., Castañeda-Rico, S., Murphy, K. R., Dobrynin, P., Ferrie, G. M., Haefele, H., Maldonado, J. E., et al., 2020. Comparison of genomic diversity and structure of sable antelope (*Hippotragus niger*) in zoos, conservation centers, and private ranches in North America. *Evolutionary Applications*, 13: 2143-2154.

Gremme, G., Steinbiss, S. and Kurtz, S., 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* 10:645–656.

Hewitt, G., 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139: 617–638.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., et al., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44: D286–D293.

Huerta-Cepas, J., Forslund, K., Coelho, L., Szklarczyk, D., Jensen, L., von Mering, C. and Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, 34: 2115–2122.

Humble, E., Dobrynin, P., Senn, H., Chuven, J., Scott, A. F., Mohr, D. W., Dudchenko, O., Omer, A. D., et al., 2020. Chromosomal-level genome assembly of the scimitar-horned oryx: insights into diversity and demography of a species extinct in the wild. *Molecular Ecology Resources*, 20:1668–1681.

IUCN SSC Antelope Specialist Group, 2017. *Hippotragus equinus*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10167A50188287.en on April 8th, 2020.

Jansen van Vuuren, B., Rushworth, I. and Montgelard, C., 2017. Phylogeography of oribi antelope in South Africa: evolutionary versus anthropogenic panmixia. *African Zoology*, 52: 189–197.

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G. and Allendorf, F. W., 2016. Genomics

advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9: 1205–1218.

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. and Frith, M. C., 2011. Adaptive seeds tame genomic sequence comparison. *Genome Research*, 21: 487–493.

Kircher, M., Sawyer, S. and Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40: e3–e3.

Koepfli, K.-P., Tamazian, G., Wildt, D., Dobrynin, P., Kim, C., Frandsen, P. B., Godinho, R., Yurchenko, A. A., et al., 2019. Whole genome sequencing and re-sequencing of the Sable antelope (*Hippotragus niger*): a resource for monitoring diversity in *ex-situ* and *in-situ* populations. *G3:Genes, Genomes, Genetics*, 9: 1785–1793.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M. A., 2009. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19:1639–1645.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25: 2078–2079.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*: 1303–3997.

Makałowski, W., Gotea, V., Pande, A. and Makałowska, I., 2019. Transposable elements: classification, identification, and their use as a tool for comparative genomics, pp. 177–207, in *Evolutionary Genomics*. Humana Press Inc, New York.

Manchanda, N., Portwood, J. L., Woodhouse, M. R., Seetharam, A. S., Lawrence-Dill, C. J., Andorf, C. M. and Hufford, M. B., 2020. GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC Genomics*, 21: 1-9.

Martchenko, D., Prewer, E., Latch, E. K., Kyle, C. J. and Shafer, A. B. A., 2018. Population genomics of ungulates, pp. 1-25, in *Population Genomics*. Springer International Publishing Inc (P. A. Hohenlohe and O. P. Rajora, eds.), Cambridge.

Matthee, C. A. and Robinson, T. J., 1999. Mitochondrial DNA population structure of roan and sable antelope; implications for the translocation and conservation of the species. *Molecular Ecology*, 8: 227–238.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20: 1297–1303.

Meyer, M. and Kircher, M., 2010. Illumina sequencing library preparation for highly

multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010: pdb.prot5448.

Norman, A. J., Putnam, A. S. and Ivy, J. A., 2019. Use of molecular data in zoo and aquarium collection management: benefits, challenges, and best practices. *Zoo Biology*, 38: 106–118.

Okonechnikov, K., Conesa, A. and García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32: 292–294.

Ou, S., Chen, J. and Jiang, N., 2018. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research*, 21: e126.

Platt, R. N., Vandewege, M. W. and Ray, D. A., 2018. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Research*, 26: 25-43.

Price, A. L., Jones, N. C. and Pevzner, P. A., 2005. De novo identification of repeat families in large genomes. *Bioinformatics*, 21.suppl_1: i351–i358.

Qi, W., Jiang, X., Yan, C., Zhang, W., Xiao, G., Yue, B. S. and Zhou, C. Q., 2018. Distribution patterns and variation analysis of simple sequence repeats in different genomic regions of bovid genomes. *Scientific Reports*, 8: 1–13.

Reese, J. T., Childers, C. P., Sundaram, J. P., Dickens, C. M., Childs, K. L., Vile, D. C. and Elsik, C. G., 2010. Bovine Genome Database: supporting community annotation and analysis of the *Bos taurus* genome. *BMC Genomics*, 11: 1–8.

Rowe, K. C., Singhal, S., Macmanes, M. D., Ayroles, J. F., Morelli, T. L., Rubidge, E. M., B,i K. and Moritz, C. C., 2011. Museum genomics: Low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources,* 11:1082–1092.

Saha, S., Bridges, S., Magbanua, Z. and Peterson, D., 2008. Empirical comparison of *ab initio* repeat finding programs. *Nucleic Acids Research*, 36: 2284–2294.

Schibler, L., Roig, A., Mahe, M. F., Laurent, P., Hayes, H., Rodolphe, F. and Cribiu, E. P., 2006. High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics*, 7: 194.

Shafer, A., Wolf, J., Alves, P., Bergström, L., Bruford, M., Brännström, I., Colling, G., Dalén, L., et al., 2015. Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30: 78–87.

Smit, A. F. A., Hubley, R. and Green, P., 2015. RepeatMasker Open-4.0. 2013-2105. Available from http://www.repeatmasker.org.

Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding.

Bioinformatics, 24: 637–644.

Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., Egholm, M., Rothberg, J. M., et al., 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences*, 103:13578–13584.

Supple, M. and Shapiro, B., 2018. Conservation of viodiversity in the genomics era. *Genome Biology*, 19: 1–12.

Swillens, S., Goffard, J. C., Marechal, Y., de Kerchove d'Exaerde, A. and El Housni, H., 2004. Instant evaluation of the absolute initial number of cDNA copies from a single real-time PCR curve. *Nucleic Acids Research*, 32:e56–e56.

van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., et al., 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43: 11–10.

van Dijk, E., Jaszczyszyn, Y., Naquin, D. and Thermes, C., 2018. The third revolution in sequencing technology. *Trends in Genetics*, 34: 666–681.

Wang, K., Wang, L., Lenstra, J., Jian, J., Yang, Y., Hu, Q. and Abbott, R., 2017. The genome sequence of the wisent (*Bison bonasus*). *GigaScience*, 6: gix016.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. and Jaffe, D. B., 2017. Direct determination of diploid genome sequences. *Genome*, 27: 757–767.

Zdobnov, E., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R., Simão, F., Ioannidis, P. , Seppey, M., Loetscher, E. and Kriventseva, E. V., 2017. OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, 45: D744–D749.

Zeng, L., Kortschak, R., Raison, J., Bertozzi, T. and Adelson, D., 2018. Superior *ab initio* identification, annotation and characterisation of TEs and segmental duplications from genome assemblies. *Plos One*, 13: e0193588.

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C. and Weir, B. S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28: 3326–3328.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., et al., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology*, 10: R42.

3.11 Supporting Information

Table S3. 11 Genome assembly statistics calculated by Supernova, using randomly selected reads to achieve 56x depth of coverage.

	Assembly run
Read number (M)	1212.03
Mean read length (bp)	139.50
Raw coverage (x)	60.70
Effective mean coverage (x)	39.70
Read two Q30 (%)	80.70
Median insert (bp)	335.00
Proper pairs (%)	90.32
Molecule length (Kbp)	21.57
Heterozygosity distance (bp)	636.00
Unbarcoded reads (%)	7.05
N50 reads per barcode (K)	0.94
Duplicates (%)	26.01
Phased (%)	48.13
Scaffolds >= 10Kbp (K)	2.20
N50 edge size (Kbp)	30.36
N50 contig size (Kbp)	239.59
N50 phased block size (Mbp)	1.40
N50 scaffold size (Mbp)	8.42
Base missing from scaffolds >= 10Kb (%)	4.10
Assembly size (scaffolds >= 10Kb) (Gbp)	2.56
Estimated genome size (Gbp)	3.01

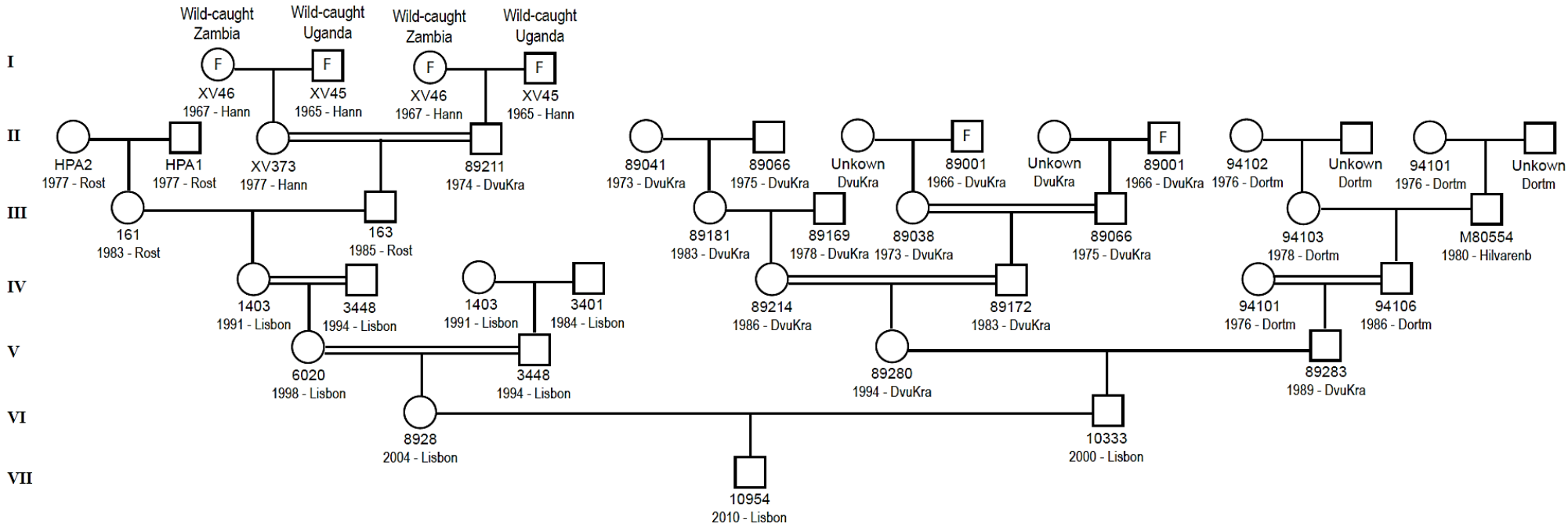
x – covered reads; K – thousand; M – million; bp – base-pairs; Kbp – thousand base-pairs; Mbp – million base-pairs; Gbp – billion base-pairs.

Table S3. 12 Genome annotation quality assessment using GenomeQC tool, comparing roan antelope with cow reference genome (BosTau9), sable antelope genome and the scimitar-horned oryx genome.

	Roan antelope	**Domestic cow**	**Sable antelope**	**Scimitar-horned oryx**
Number of gene models	30,622	30,589	23,846	28,559
Minimum gene length	56	45	6	6
Maximum gene length	1,118,330	2,433,917	163,506	162,991
Average gene length	37,281.0	39,806.5	16,709.8	16,128.4
Number of transcripts	25,139	66,643	23,846	28,559
Average number of transcripts per gene model	2.0	2.2	1.0	1.0

Table S3. 13 Alignment statistics of the roan antelope genome against the domestic cow reference genome (BosTau9), per chromosome.

	Identity (%)	Mapped bp (N)	Mapped scaffolds (N)	Mapping (%)
chr1	92.89	155,462,201	32	98.06
chr2	93.03	134,758,668	41	98.92
chr3	93.00	116,791,714	64	96.52
chr4	92.91	116,204,881	28	96.84
chr5	92.88	114,968,136	52	95.74
chr6	92.81	115,955,195	24	98.43
chr7	92.92	106,359,928	68	96.09
chr8	92.97	110,519,018	43	97,53
chr9	92.82	102,494,769	16	97.19
chr10	92.99	98,360,627	29	95.21
chr11	92.83	105,507,747	26	98.62
chr12	92.51	83,334,402	23	95.55
chr13	92.84	79,717,752	32	95.50
chr14	92.85	80,159,661	21	97.28
chr15	92.77	70,396,358	44	82.81
chr16	92.72	77,921,579	39	96.18
chr17	92.55	71,298,713	28	97.45
chr18	92.65	61,678,633	58	93.71
chr19	92.49	60,956,150	40	96.07
chr20	92.61	69,947,538	8	97.18
chr21	92.65	67,359,392	61	96.42
chr22	92.50	59,787,467	21	98.38
chr23	92.45	46,624,102	38	88.81
chr24	92.62	61,424,422	16	98.57
chr25	92.14	41,507,042	19	98.01
chr26	92.54	49,906,710	27	95.99
chr27	92.37	43,147,230	12	94.60
chr28	92.55	44,358,876	29	96.56
chr29	92.25	48,728,272	27	95.36
chrX	93.30	115,206,517	261	82.88



Lisbon - Lisbon Zoo (Portugal)
DvuKra - Safari Park Dvur Králové (Czech Republic)
Dortm - Dortmund Zoo (Germany)
Rost - Rostock Zoo (Germany)
Hilvarenb - Hilvarenbreek Zoo (Netherlands)
Hann - Hannover Zoo (Germany)

F - Zoo Founder

Figure S3. 7 Pedigree of roan antelope individual ID:10954, used to produce the assembled genome. Six generations are depicted, indicated by Roman numbers, with generation I made up of wild-caught individuals to become zoo founders (F). Males are indicated by squares and females by circles. Consanguineous mattings are indicated by double line connections between male and female. Each individual is identified by an ID code, year of birth and zoo name. Adapted from pedigree made available by the Lisbon Zoological Garden.

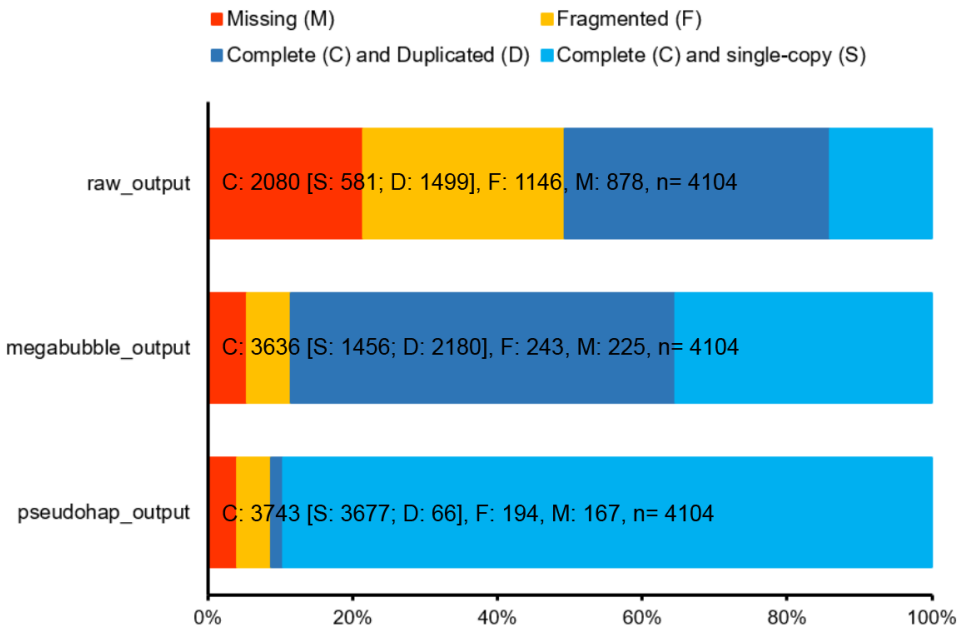


Figure S3. 8 Plot comparing gene completeness of the roan genome assembly according to BUSCO percentage scores, among the three Supernova assembly output styles: raw; megabubble; and pseudohaplotype. BUSCO scores were estimated using the Mammalia OrthoDB v9 gene set, containing 4,104 genes. Total number of genes falling into the four categories, according to the legend and colour code.

CHAPTER 4

Comparative genomics within *Hippotragus* spp.

Comparative population genomics of the roan and the sable antelope: population structure and drivers of species diversity

Margarida Gonçalves[1,2], Margarida Andrade[1,2], Hans R. Siegismund[3], Bettine Jansen van Vuuren[4], Nuno Ferrand[1,2,4], and Raquel Godinho[1,2,4]

Author's affiliations: [1]CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus de Vairão, Portugal; [2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto; [3]Department of Biology, University of Copenhagen, Denmark; [4]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South-Africa

4.1 Introduction

Comparative population genomics between species can elucidate how differential patterns of genetic diversity are shaped by both intrinsic – biological and behavioural – and extrinsic – ecological – factors (Coleman, Chisholm and Karl 2010; Romiguier et al. 2014). Today, we know that the diversity of a species is firstly determined by its evolutionary strategy (Romiguier et al. 2014), however, other contributing factors, including population history, certainly play important roles (Chamaillé-Jammes et al. 2008; Wang et al. 2009; Leffler et al. 2012). In terms of evolutionary strategy, ungulates are characterized as density-dependent K-strategists, whose population sizes are dependent on environmental carrying capacity (McCullough 1999; Bonenfant et al. 2009). Because of this, they typically have long life spans and generation times, low adult mortality enabling multiple reproductive efforts over a lifetime, as well as small litter size, with high maternal investment in young (Bowyer et al. 2014). Shared life-history characteristics increases interspecific competition among large herbivores, reducing the environmental carrying capacity (Stewart et al. 2002; Kentie et al. 2020). Ecological specialization, on the other hand, facilitates closely related species to co-occur through niche partitioning (Weber and Strauss 2016; Estevo, Nagy-Reis and Nichols 2017; Zheng et al. 2020). Across continental Africa, the distribution and ecological specialization of large herbivores, are mostly explained by climatic variability, biome density, and geographical area, however, with a strong latitudinal variation and hemispheric asymmetry (Fernández and Vrba 2005; Zhang, Zhang and Ma 2016). At the regional scale, habitat heterogeneity promotes specialization (Pontarp, Ripa and Lundberg 2015).

Two of the most iconic and emblematic African antelope species are the roan (*Hippotragus equinus*) and the sable (*H. niger*), the only extant representatives of the genus. Both species have diverged around 5.8 Mya (Bibi 2013) and still share many life-history characteristics and ecological conditions, thus representing powerful models for comparative genomics. As large herbivores, they have long generation times and high maternal investment in young, living in breeding herds with a strong social structure (Chardonnet and Crosmary 2013; Estes 2013). Both inhabit sub-Saharan Africa mesic woodland savannas, and are found in sympatry throughout most of their distributional range, which is split in several recognized subspecies, six for the roan, and four for the sable antelope (IUCN SSC Antelope Specialist Group 2017a, 2017b). However, the roan antelope presents a wider distributional range, covering both hemispheres, across several biogeographical regions and biomes (White 1983; Linder et al. 2012). The sable antelope, on the contrary, is only distributed across the Southern hemisphere, strongly associated with the miombo woodland biome (Estes 2013), therefore being considered as a more specialized species.

Both species have been the focus of several studies assessing their levels of population diversity, structure, and differentiation at the intraspecific level, using traditional molecular markers, such as mitochondrial DNA and nuclear microsatellites. For the roan antelope, a total of five nuclear groups and four mitochondrial lineages were recently described using a comprehensive sampling through its whole distributional range (Gonçalves et al. this thesis). These authors have extended the evaluation of genetic diversity of each population, after previous work by Matthee and Robinson (1999) and Alpers et al. (2004), that initiated the description of such parameters. For the sable antelope, the same structural partitioning of five nuclear groups and four mitochondrial lineages was described, using the same molecular markers, and covering, as well, the whole species range (Rocha 2014; Vaz Pinto 2018). Likewise, these authors have expanded previous work on the species genetic diversity by Matthee and Robinson (1999), Pitra et al. (2002, 2006), and Jansen van Vuuren et al. (2010). Throughout these studies, the evolutionary history of both species was consistently associated with past climatic and geomorphological events that promoted vicariance in many sub-Saharan African taxa, namely ungulates (Lorenzen, Heller and Siegismund 2012).

Recently, whole-genome data has provided further evidence for shared evolution and adaptive traits across ungulate species and other ruminants (Chen et al. 2019). The recent release of several African antelope reference genomes, including the one for roan (Gonçalves et al. this thesis) and sable antelope (Koepfli et al. 2019), enabled us to understand the evolutionary history of these species through a genomic perspective. The application of genome-wide data across multiple individuals, excluding non-neutral loci,

allows the inference of population genetic parameters and phylogenetic history of natural populations with deeper resolution (Black IV et al. 2001; Luikart et al. 2003), a fact that has been consistently observed across several species of mammals (e.g. Demos et al. 2015; Sinding et al. 2018; Ayoola et al. 2020). Ideal research on population genomics typically includes dozens of individuals, representative of the geographical distribution of the species, with each individual sequenced at high depth of coverage (>20x), to determine genotypes (Luikart et al. 2003; Ellegren 2014). However, this sequencing design can generate methodological and economic constrains. Therefore, a genotype likelihood framework allows for cost-effective studies of many individuals sequenced at lower coverage, bypassing the need of individual genotyping (Kim et al. 2011; Nielsen et al. 2011). Recently developed software (Korneliussen, Albrechtsen and Nielsen 2014) uses genotype likelihoods to infer the distribution of allele frequencies across populations (Nielsen et al. 2012), allowing for comparative measures at the intraspecific level.

In this study we provide the first comparative population genomic analysis for the roan and sable antelopes. Using whole genome sequencing data and covering both species' native distributional range, we follow a genotype likelihood framework to provide insight into population diversity, structure, and differentiation, as well as the phylogenetic relationships between the different populations of both species. We aim to compare the genetic patterns obtained for the roan and sable antelopes, and interpret this comparison from a biological and ecological perspective, relating life-history characteristics with range distribution and habitat specialization across African savannas. We also highlight the advantages of using whole-genome data, comparing the obtained results with previous studies, namely on nuclear data.

4.2 Materials and Methods

4.2.1 Sample collection and DNA extraction

We used a total of 23 roan and 38 sable antelope samples, covering the total distribution range for both species (Figures 4.1 and 4.2; Table S4.1). Of these, 50 were contemporary muscle and skin samples (12 from roan and 38 from sable antelope) for which total genomic DNA was extracted using the Qiagen DNeasy Blood &Tissue Kit (Qiagen, Germany) following the manufacturer's instructions. The remaining roan antelope samples comprised historical tissues (skin, bone, and tooth) donated by Museum collections, for which we followed the Dabney et al. (2013) protocol for DNA extraction, performed in a laboratory dedicated to low-quality DNA work (see Chapter 3 – Text S3.1 for details). For both extraction methods, blank controls were routinely

included to test for contamination. Eluted DNA concentration was quantified by fluorometry, using the Qubit dsDNA HS - High Sensitivity - Assay Kit (Thermo Fisher Scientific, USA). Final DNA concentration for roan antelope samples averaged 64 ng/µl (between 10 and 110 ng/µl) and for sable antelope samples averaged 80 ng/µl (between 6 and 300 ng/µl).

4.2.2 DNA Library preparation and sequencing

Prior to library preparation, DNA of contemporary samples was sheared by sonication and fragments of ca. 350 bp were selected using AMPure XP bead clean-up protocol (Beckman Coulter, USA). Double stranded DNA library preparation followed Meyer and Kircher (2010) and Kircher and co-workers (2012) protocols for dual-indexing with regular Illumina adaptors. Amplification and purification followed Dabney and Meyer (2012), after determining by qPCR the number of amplification cycles that would minimize hydrolytic damage (Swillens et al. 2004; Stiller et al. 2006) (Table S4.1). Amplified libraries from both contemporary and historic samples were cleaned using MinElute PCR purification kit (Qiagen, Germany), quantified by Kappa Library Quantification Kit (Roche Sequencing and Life Science, USA), and pooled at equimolar amounts. Pooled sequencing was performed on a HiSeq 4000 platform (Illumina, USA) using the 150 bp paired-end sequencing protocol.

4.2.3 Read mapping and filtering

Samples were demultiplexed based on their unique indexes and adaptors were removed using Trimmomatic v.0.36 (Bolger, Lohse and Usadel 2014). Minimum read length was set to 50 bp for contemporary samples, and 30 bp for historical samples. After quality control with FastQC v.0.11.8 (Andrew 2010) and MultiQC v.1.8 (Ewels et al. 2016), reads were mapped against sable antelope reference genome (Genbank accession number: GCA_006942125.1; Koepfli et al. 2019), using BWA-mem with default settings (Li 2013). After mapping, duplicated reads were marked with Picard v.2.21.4 (http://broadinstitute.github.io/picard/) and local realignment around indels was improved using GATK v.3.8 IndelRealigner (McKenna et al. 2010). Final mapping quality was assessed with QualiMap v.2.2.2 (Okonechnikov, Conesa and García-Alcalde 2016). Raw variant filtering was performed with BCFtools v.1.9 (Li et al. 2009) and snpCleaner (https://github.com/tplinderoth/ngsQC/tree/master/snpCleaner), discarding indels and marked duplicates. We only considered sites in which at least 80% of individuals had ≥

1x coverage. We increased sensibility to detect heterozygotes (-h 0; -H $1x10^{-4}$), as well as quality base biases (-b $1x10^{-20}$; -f $1x10^{-4}$), according to author's recommendations.

4.2.4 Genotype likelihood and SNP calling

To accurately incorporate the statistical uncertainty associated with low-coverage data, analyses were conducted in a genotype likelihood framework implemented in ANGSD (Analysis of Next Generation Sequencing Data) software v.0.930 (Nielsen et al. 2011; Korneliussen et al. 2014). Using previously filtered variant-sites and analysing data within species, genotype likelihoods were calculated by removing reads with i) a flag equal or above 256 (-remove_bads 1), ii) multiple hits (-uniqueOnly 1), iii) mapping quality below 20 (-minMapQ 20), and iv) reads whose mate could not be mapped (-only_proper_pairs 1). We only considered sites below 3-fold the average coverage across individuals. Base quality score was filtered above 20 (-minQ 20) and only genotypes with a likelihood ratio test statistic below 24 (-SNP_pval $1x10^{-6}$) were kept, using the samtools genotype model (-GL 1). The distribution of sample allele frequencies was estimated through site-frequency spectra across all sites and individuals, with no need to call for genotypes (Nielsen et al. 2012). Per-site allele frequency was estimated by fixing the major allele frequency, using the sable antelope reference genome as the ancestral state, and calculating the minor allele frequency based upon genotype likelihoods.

4.2.5 Population structure

Within species, and using genotype likelihoods, population structure was analysed by both Principal Component Analysis (PCA) and admixture proportions. PCAngsd v.0.985 (Fumagalli et al. 2013) was used to estimate covariance matrices and plot the PCAs under default parameters. NGSadmix (Skotte, Korneliussen and Albrechtsen 2013) was used to estimate individual genome admixture proportions, testing for K = 1 to K = 8, with 10 runs per K. Best K was then checked for in CLUMPAK server (Kopelman et al. 2015). For both analyses, minimum allele frequency was filtered for 1% (-minMaf 0.01). According to the population structure obtained, we also performed a per-population multi-sample genotype likelihood estimate (-doSaf 1), using the folded site-frequency spectra with realSFS as implemented on ANGSD (Nielsen et al. 2012). Isolation by distance within population was checked for with Mantel's test (Mantel 1967; Slatkin 1993), where the null hypothesis assumes no significant correlation between genetic and geographical distance. Individual pairwise genetic distances were obtained with ANGSD, after

computing genotype posterior probabilities for each of the three possible genotypes at each locus (-doGeno 8), using the estimated allele frequency as prior to retain SNPs. This strategy provides the most accurate estimate of genetic distances with no need for genotype calling (Vieira et al. 2016). Pairwise Euclidean distances (Diniz-Filho et al. 2013) between sampling-sites were used as geographical distances; a distance = 0 was assumed for individuals within the same sampling-site. Mantel's tests were run using the function “mantel” from the R package *vegan* (Oksanen et al. 2019), with the Pearson correlation model (Diniz-Filho et al. 2013), 10,000 permutations, and checking for significance with p-value < 0.05.

4.2.6 Diversity and differentiation

Relatedness between individuals was calculated for each species using population allele frequencies and estimating kinship coefficients with NgsRelate v.2 (Korneliussen and Moltke 2015), described by values of the three Jacquard coefficients (F_a, F_b, and theta). Hardy-Weinberg equilibrium was tested for each population for both species, using population folded site-frequency spectra (-doHWE 1).

Within-population pairwise linkage disequilibrium (LD) was estimated with ngsLD (Fox et al. 2019). Since this statistic is sensible to low coverage, we compared two different datasets per population: 1) considering all individuals; 2) considering only the 3 individuals with the highest coverage (roan antelope: 4.5x – 14.3x; sable antelope: 8.8x – 15.7x). Linkage decay rate over physical distance was assessed by fitting the distribution of randomly sampled 50,000 SNPs per population using the r^2 model (Hill and Weir 1988; Remington et al. 2001) across 100 bootstrap replicates. Plots were attained with the incorporated fit_LDdecay script. Dataset 2 achieved a better fit between the data and the model (Figure S4.1 and Figure S4.2) and was used to set the sliding-window size for the subsequent analyses.

Genome-wide heterozygosity was estimated per-sample using the folded site-frequency spectrum obtained with realSFS and 1,000 bootstrap replicates. Nucleotide diversity was estimated from population folded site-frequency spectra using the full maximum likelihood model (-doThetas 1) and thetaStat, as implemented in ANGSD (Korneliussen et al. 2013). Both Watterson theta estimator (Watterson 1975) and Tajima’s D (Tajima 1989) values were assessed for within population comparisons across sliding windows of 100,000 bp, spaced every 25,000 bp. Population pairwise fixation index (F_{ST}) was calculated from two-populational folded site-frequency spectra, averaged across 1,000 bootstrap replicates; sliding-window analyses were set as

described above. Unweighted statistics and respective standard-deviations were compared for all pairs of populations, within species.

4.2.7 Phylogenetic reconstruction

Phylogenetic relationships within species were reconstructed using individual pairwise genetic distances, obtained in ANGSD, after computing genotype posterior probabilities, as previously described. As outgroups for the roan antelope, we used a sable antelope sample from Kenya (Hn325) and another from Zimbabwe (Hn327); as outgroup for the sable antelope, we used two roan antelope samples, one from Ghana (He80) and another from Tanzania (He95). Posterior probabilities were then used in ngsDist v.1.0.8 (Vieira et al. 2016) to obtain the pairwise distance matrix between individuals of each species, in a block-bootstrapping procedure of 100 distance matrices, obtained from repetitively sampling blocks of 500 SNPs. FastME v.2.1.5 (Lefort, Desper and Gascuel 2015) was used to reconstruct neighbour-joining trees for each species matrix, with the SPR tree topology improvement (-s). RAxML-NG v.0.9 (Kozlov et al. 2019) was used to plot the main tree with bootstrap node support, which was then visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

4.3 Results

4.3.1 Whole-genome data

We generated whole-genome sequencing data from 23 roan and 38 sable antelope samples, representative of both species distributional range (Figures 4.1 and 4.2; Table S4.1). After quality control, alignment to the sable antelope reference genome (Koepfli et al. 2019) resulted in an average 99.3% of mapped reads for the roan antelope, except for two samples (He29 – 52% and HeNI124 – 77.3%), and 99.7% for the sable antelope. Average base mapping quality was above 20: between 19.4 and 26.6 for roan, and 27.7 and 31.9 for sable antelope. Average depth of coverage across species ranged between 5x and 10x: for roan antelope, individual depth achieved ranged between 1.5x and 14.3x; for sable antelope, was between 2.3x and 15.7x. A total of 11,735 contigs analysed for roan and 12,140 contigs analysed for sable antelope, allowed to retrieve 33,563,489 SNPs covering 80% of roan antelope individuals and 28,303,928 SNPs covering 80% of sable antelope individuals.

4.3.2 Population structure

For the roan antelope structure analyses, we found two clusters (Figure 4.1). PCA showed a clear correlation with the geographical distribution of samples (Figure 4.1a). PC1, with an explained variance of 52.7%, separated samples from a Northern cluster (Guinea-Bissau, Ghana, Cameroon, Chad, and Central African Republic) (Fig. 4.1 – sampling localities GB1, GB2, Gh1, Gh2, Cm1, Cm2, Ch and CAR) from samples of a Southern cluster (Ethiopia, Uganda, Tanzania, Zambia, South-Africa, Namibia, and Angola) (Fig. 4.1 – sampling localities Et1, Et2, Ug, Ta1-Ta5, Za1, Za2, SA, Na, and An1-An3). PC2, with an explained variance of 6.4%, retrieved the dispersion of Northern individuals along the geographical axes west-east. Mean likelihood values [ln(K)] of admixture analyses showed an increase from K = 2 onwards (Figure S4.3a; Figure S4.4), however with a clear maximum in delta (K) statistics for K = 2 (Figure S4.3b). For two clusters, Northern individuals separated from those in the South, as portrayed in the PCA, with signatures of admixture between the two clusters exhibited by the individual from Central African Republic and the two individuals from Ethiopia (Fig. 4.1 – sampling localities CAR, and Et1 and Et2, respectively). Checking for isolation by distance within population, we found significant ($p < 0.05$), albeit not strong ($r \approx 0.4$), positive correlation for both Northern and Southern clusters (Table S4.2).

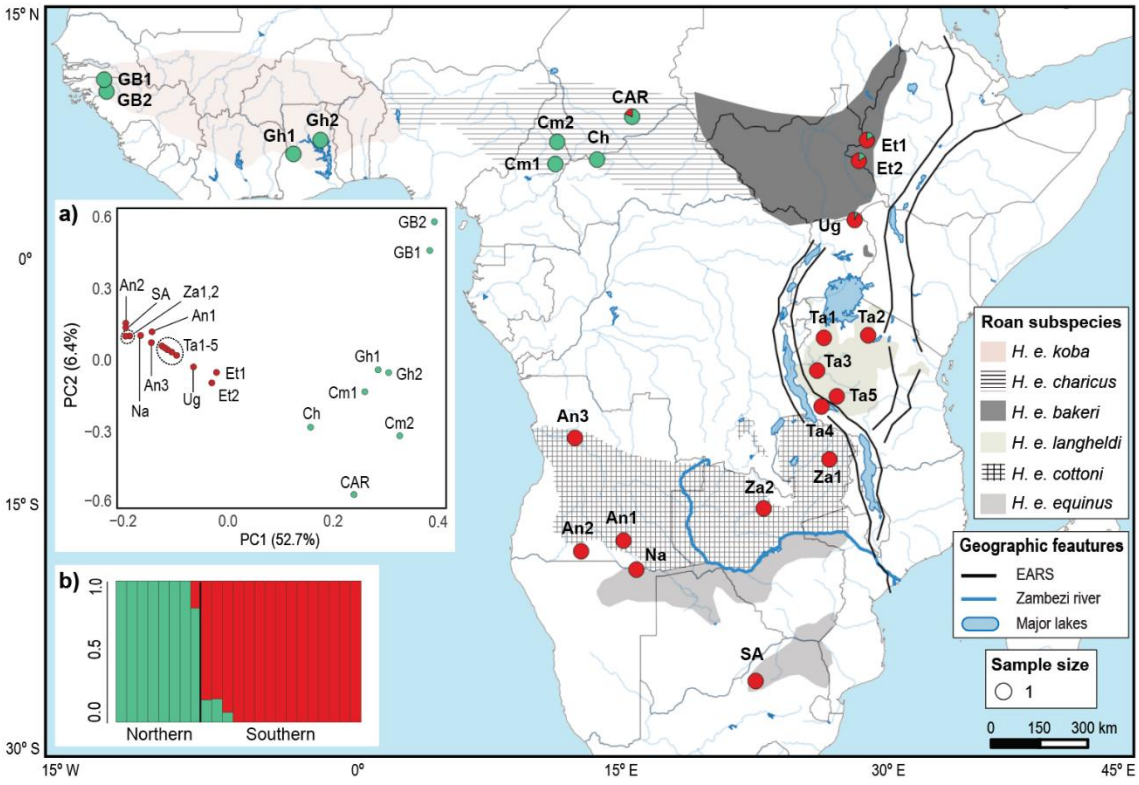


Figure 4. 1 Population structure of roan antelope (*H. equinus*). The map shows the geographical distribution of assigned clusters, according to the admixture analysis performed with NGSadmix, for K = 2. Each sampling location identified by a different number, and according to country, as follows: GB – Guinea-Bissau; Gh – Ghana; Cm – Cameroon; Ch – Chad;

CAR – Central African Republic; Et – Ethiopia; Ug – Uganda; Ta – Tanzania; Za – Zambia; An – Angola; Na – Namibia; and SA – South-Africa (Table S4.1). Pie-charts represent the average assignment probability to each cluster, within each location: northern in green; southern in red. Species distribution retrieved from IUCN (2017a) and subspecies limits adapted from Ansell (1972). Also depicted are geographic features mentioned in the text, namely the EARS in black, and the Zambezi River in dark blue. **a)** Scatterplot of the first two principal component analysis with PCAngsd, showing the percent of explained variance of each component (PC). Each colour represents a different cluster and sampling-codes as in the map. **b)** Representation of individual assignment probability ($0.0 < q < 1.0$) to each cluster. Each vertical bar represents an individual, coloured according to assigned cluster.

For sable antelope, structure analyses were able to retrieve five well-defined clusters (Figure 4.2). As for roan, PCA on sable antelope showed a correlation with the geographical distribution of samples (Figure 4.2a). The first axis (PC1), explaining 37.7% of the variance, separated Angolan individuals from the remaining ones (Fig. 4.2 – sampling localities An1 and An2), while the second axis (PC2), explaining 18.0% of the variance, separated west Tanzanian individuals from the remaining individuals (Fig. 4.2 – sampling localities Ta1-Ta5). Admixture analyses clearly support five clusters (Figure 2b). Mean likelihood [ln(K)] showed a gradual increase from K = 4 to K = 8 (Figure S4.5a; Figure S4.6), with K = 5 showing the lowest standard-deviation and highest delta (K) value after K = 2 (Figure S4.5b). The five clusters comprise the individuals from i) Angola, ii) west Tanzania, iii) Kenya, south-eastern Tanzania, and northern and central Mozambique, defining the Eastern cluster, iv) Zambia, and v) southern Mozambique, Zimbabwe, South-Africa, Botswana, and Namibia, defining the Southern cluster. Signatures of admixture were observed between Eastern and Southern clusters. Checking for isolation by distance within population, we only found significant positive correlation ($p < 0.05$) for the Eastern and Southern clusters, with a stronger correlation within the first (Eastern $r \approx 0.8$; Southern $r \approx 0.5$) (Table S4.2).

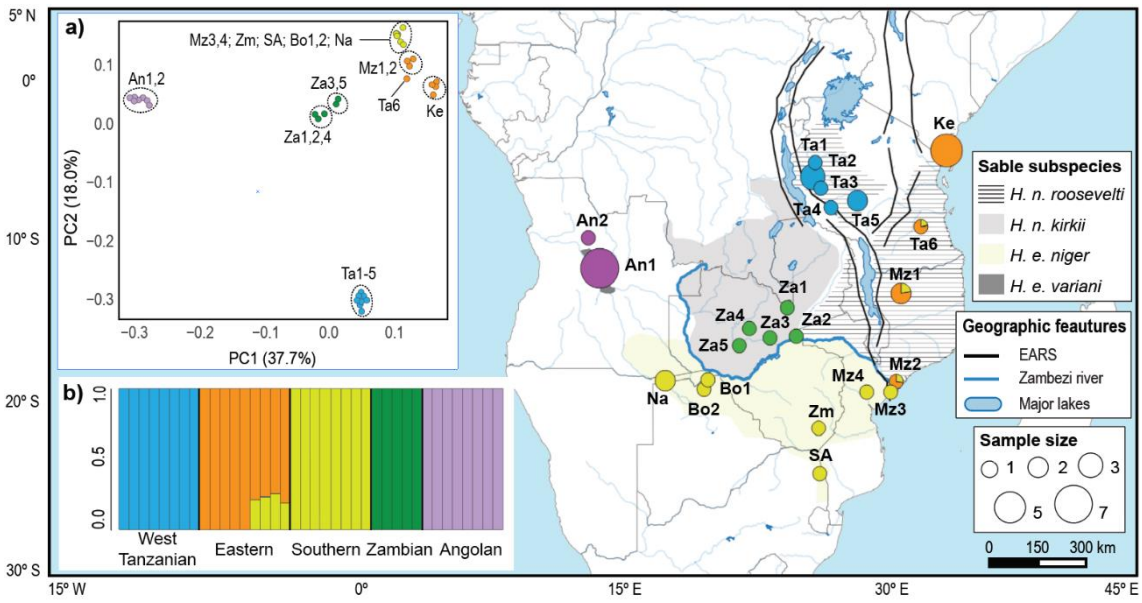


Figure 4. 2 Population structure of sable antelope (*H. niger*). The map shows the geographical distribution of assigned clusters, according to the admixture analysis performed with NGSadmix, for K = 5. Each sampling location identified by a different number, and according to country, as follows: Ta – Tanzania; Ke – Kenya; Mz – Mozambique; Za – Zambia; An – Angola; Na – Namibia; Bo – Botswana; Zm – Zimbabwe; and SA – South-Africa (Table S4.1). Pie-charts represent the average assignment probability to each cluster, within each location: west tanzanian in blue; eastern in orange; southern in yellow; zambian in green; angolan in purple. Species distribution retrieved from IUCN (2017b) and subspecies limits adapted from Ansell (1972). Also depicted are geographic features mentioned in the text, namely the EARS in black, and the Zambezi River in dark blue. **a)** Scatterplot of the first two principal component analysis with PCAngsd, showing the percent of explained variance of each component (PC). Each colour represents a different cluster and sampling-codes as in the map. **b)** Representation of individual assignment probability ($0.0 < q < 1.0$) to each cluster. Each vertical bar represents an individual, coloured according to assigned cluster.

4.3.3 Diversity and differentiation

Based on the estimated kinship coefficients within populations of each species, all individuals were considered unrelated (results not shown). The ratio of site allele frequencies was similar considering or not considering HWE (Table S4.3). Therefore, populations were assumed to be in equilibrium.

For roan antelope, pairwise LD decay rate was $3.4x10^{-4}$ for the Northern cluster and $1.5x10^{-4}$ for the Southern cluster. Both showed overlaid maximum values of LD (Northern LDmax = 0.46; Southern LDmax = 0.47), with Northern showing higher values of minimum LD (Northern LDmin = 0.37; Southern LDmin = 0.29). Both LD decay curves had high asymptotes, which may indicate population structure or small sample size (Figure S4.2). Genome-wide heterozygosity indicated a higher proportion of heterozygous SNPs for Northern individuals (average of 0.129), compared to the Southern cluster (average of 0.116) (Figure 4.3a). Lowest and highest heterozygosity were recorded for a Zambian individual (sampling locality Za1; $H_e = 0.058$) and the Ugandan individual (sampling locality Ug; $H_e = 0.191$), respectively. Nucleotide diversity, estimated using Watterson's theta (Watterson 1975), as well as proportion of allele frequencies per population, suggested the same trend observed for the heterozygosity (Figure 4.3b; Figure S4.7a). Overall negative mean values of Tajima's D (Tajima 1989) for both clusters (Northern = $-1.14x10^{-6}$; Southern = $-1.08x10^{-6}$) most likely reflects population structure (Table S4.4), which was visible in the PCA and admixture plot (Figure 4.1; Figure S4.4).

For sable antelope, pairwise linkage decay rate was between $8.9x10^{-5}$ for West Tanzanian cluster, and $1.2x10^{-4}$ for Southern cluster. Highest values of LD were attained for the Angolan cluster ($0.64 < LD < 0.42$), whereas Eastern had the lowest values ($0.49 < LD < 0.31$). The shape of the LD decay curve for Angolan, when compared to the remaining clusters, indicated a high intersect and possible low value of effective population size or recent bottleneck (Figure S4.2). Genome-wide heterozygosity also

indicated a decrease in the proportion of heterozygous SNPs in Angolan individuals (average of 0.093) when compared with the remaining clusters (average between 0.130 for Southern and 0.155 for West Tanzanian) (Figure 4.4a). Lowest proportion was registered for an Angolan individual (sampling locality An1; H_e = 0.073); highest proportion for a West Tanzanian individual (sampling locality Ta2; H_e = 0.177). The same trend was observed for nucleotide diversity and proportion of allele frequencies, with a decrease in Angolan's theta value (Figure 4.4b) and smaller proportion of singletons and doubletons, when compared to the remaining clusters (Figure S4.7b). Estimated levels of Tajima's D retrieved the highest mean value for Angolan cluster (mean = $7.15x10^{-7}$) and the lowest, and only negative, value for Eastern (mean = $-4.50x10^{-9}$) (Table S4.4). Overall positive values might indicate balancing selection or recent population bottleneck; the mean negative value registered for Eastern cluster most likely indicates possible population structure, which was visible in the PCA and admixture plot (Figure 4.2; Figure S4.6).

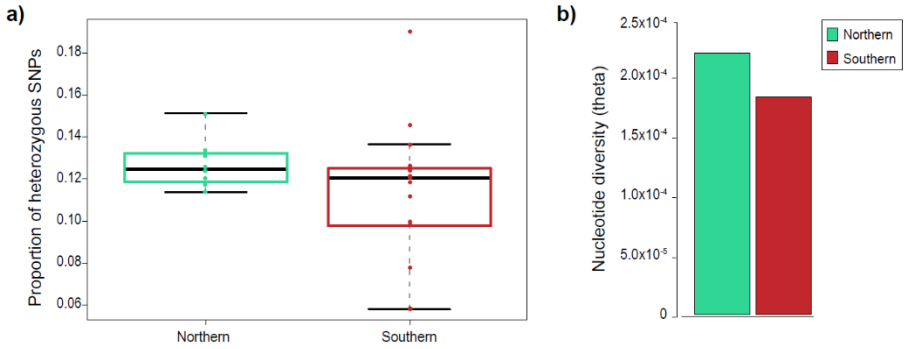


Figure 4. 3 Diversity within roan antelope (*H. equinus*) population clusters. **a)** Boxplot showing the proportion of heterozygous SNPs for all individuals, per identified cluster; **b)** Nucleotide diversity using Watterson's theta estimate (Watterson, 1975) across the same clusters, using sliding windows of 100,000 bp, spaced every 25,000 bp. Colours according to figure legend.

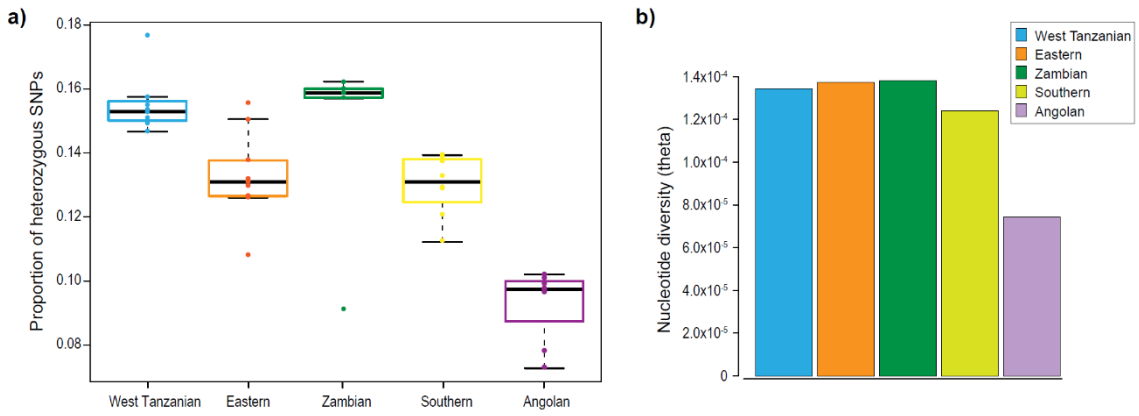


Figure 4. 4 Diversity within sable antelope (*H. niger*) population clusters. **a)** Boxplot showing the proportion of heterozygous SNPs for all individuals, per cluster. **b)** Nucleotide diversity using Watterson's theta estimate (Watterson,

1975) across the same clusters, using sliding windows of 100,000 bp, spaced every 25,000 bp. Colours according to figure legend.

Levels of F_{ST} supported the observed population structure. Within roan antelopes, value of F_{ST} between Northern and Southern cluster was of 0.338 (±0.046 – Table 4.1). Within sable antelopes, F_{ST} averaged 0.405 (±0.075 – Table 4.1). Highest values were reported between the Angolan and remaining clusters ($0.446 < F_{ST} < 0.478$), and lowest values were obtained between Zambian and remaining clusters ($0.335 < F_{ST} < 0.446$).

Table 4. 1 Genetic differentiation between clusters of roan (on top) and sable (at bottom) antelope, measured by unweighted pairwise fixation index (F_{ST}), from two-populational folded site frequency spectra, averaged across 1,000 bootstrap and using sliding-windows of 100,000 bp, spaced every 25,000 bp.

Roan (*H. equinus*)

	Northern
Southern	0.338 ± 0.046

Sable (*H. niger*)

	West Tanzanian	Eastern	Zambian	Southern
West Tanzanian				
Eastern	0.372 ± 0.071			
Zambian	0.355 ± 0.075	0.372 ± 0.064		
Southern	0.399 ± 0.068	0.355 ± 0.062	0.335 ± 0.079	
Angolan	0.464 ± 0.079	0.469 ± 0.069	0.446 ± 0.099	0.478 ± 0.079

Unweighted mean ± standard deviation across 1,000 bootstraps.

4.3.4 Phylogenetic reconstruction

Phylogenies based on within-species pairwise genetic distances retrieved the same clusters described by analyses of population structure, with highly supported bootstrap values (> 95%). For roan antelope, the phylogenetic tree divided two well-supported clades, corresponding to the Northern and Southern clusters (Figure 4.5). The only exception is the individual from Chad (sampling locality Ch), which is placed within the Southern cluster (Figure 4.1). Similarly, the Ta2 individual was placed closer to the Ugandan sample, rather than among the remaining Tanzanian samples. For sable antelope, the tree topology exhibits a first well-supported branch, separating West Tanzanian individuals from the remaining ones, which is then split into two other branches, separating Eastern and Southern, from Zambian and Angolan individuals. Each of these populations is then split into their own well-supported branch (Figure 4.6). This analysis suggests that Eastern and Southern clusters, as well as Zambian and Angolan clusters, are sister populations. Additionally, the phylogenetic analysis for sable

also retrieved subdivision within Eastern, separating Kenyan (sampling locality Ke) from the remaining samples within the cluster, a result also retrieved for K = 6 in the admixture analysis (Figure S4.6).

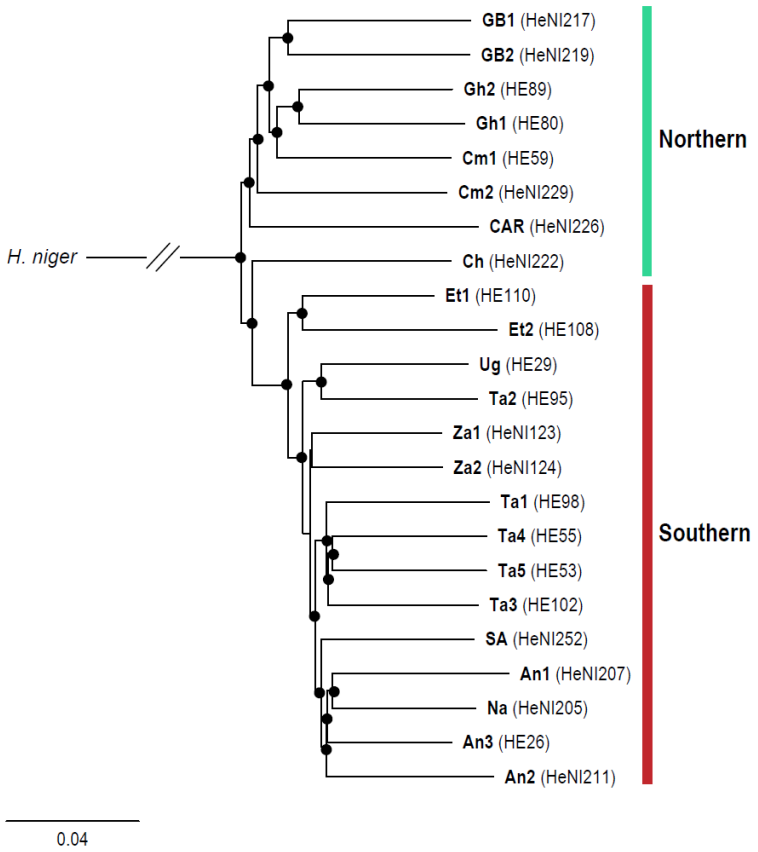


Figure 4. 5 Neighbour-joining phylogenetic tree of roan antelope (*H. equinus*), using pairwise genetic distances between individuals, measured using ngsDist, and according to represented scale, at the bottom. Tree is rooted using roan's sister-species, sable antelope (*H. niger*), and filled circles indicate bootstrap support > 95%. At the tips, individuals are identified according to sampling-code (as in Figure 4.1 and Table S4.1) and sample ID (as in Tables S4.1). Clusters identified according to Figure 4.1 and Figure 4.3.

4.4 Discussion

In this study we address the first comparative genomic study within the genus *Hippotragus*, using a comprehensive sampling of its extant species, the roan and the sable, two of the most emblematic African antelopes. Using a low- to medium-depth of sequencing coverage (1.5 – 14.3x), and following a genotype likelihood framework, we infer population genetic parameters to elucidate general patterns within species, and to compare such patterns between species. Such approach allowed us to refine our understanding of how life-history characteristics – biological factors – as well habitat specialization – ecological factors – may have contributed to differences in the genetic diversity of species and range distribution across African savannas.

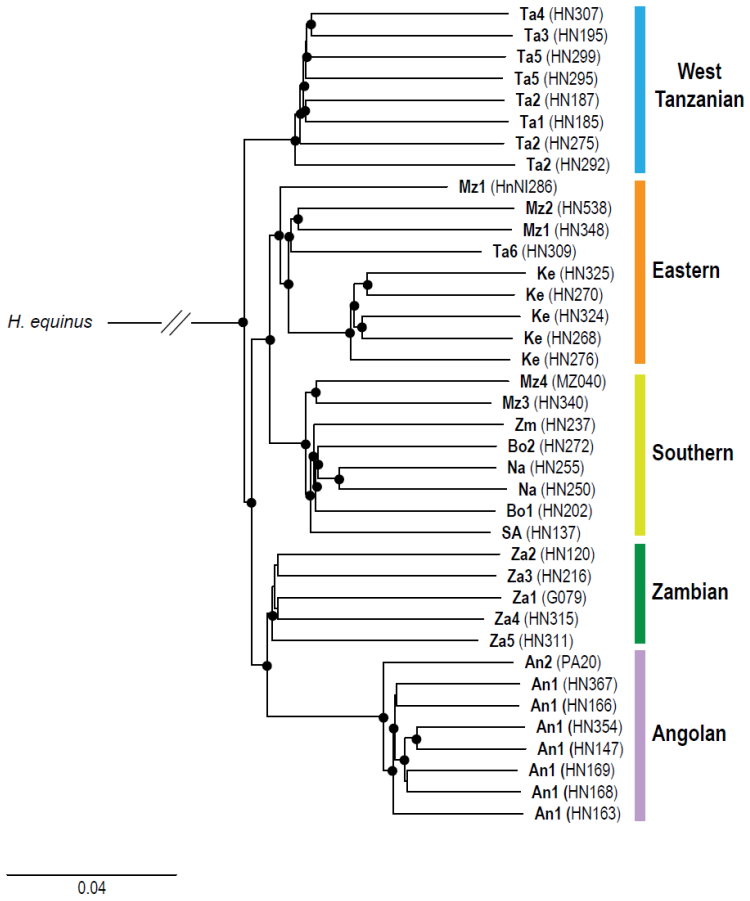


Figure 4. 6 Neighbour-joining phylogenetic tree of sable antelope (*H. niger*), using pairwise genetic distances between individuals, measured using ngsDist, and according to represented scale, at the bottom. Tree is rooted using sable's sister-species, roan antelope (*H. equinus*), and filled circles indicate bootstrap support > 95%. At the tips, individuals are identified according to sampling-code (as in Figure 4.2 and Table S4.1) and sample ID (as in Tables S4.1). Clusters identified according to Figure 4.2 and Figure 4.4.

4.4.1 Genomic insights at the intraspecific level

Roan antelope (*Hippotragus equinus*)

Genomic data supports two geographically well-defined clusters, separating individuals present in the northwest and north-central range, from those in the northeast, eastern, and southern range of the species (IUCN SSC Antelope Specialist Group 2017a). Both clusters present signs of isolation by distance, and for the northern cluster there is a considerable differentiation of individuals along the west-east geographical axis, from Guinea-Bissau towards Ghana, Cameroon, Chad, and Central African Republic. Although the split of the species in two main clusters has been previously disclosed by Alpers et al (2004) and Gonçalves et al. (this thesis), such perception of the distribution of diversity has never been possible to characterize, even using robust nuclear information from classical molecular markers. Nonetheless, some level of

population differentiation using microsatellite loci was seen before between individuals from Senegal and Cameroon (Alpers et al. 2004), and between individuals from Ghana and the remaining sites (Gonçalves et al. this thesis). The genetic partition between Northern and Southern Africa has been observed across many savanna species with similar distributional ranges to roan antelope, including ungulates (Lorenzen et al. 2012; Smitz et al. 2013; Winter, Fennessy and Janke 2018), other mammals (Bertola et al. 2016), and even birds (Miller et al. 2011). This division is coincident with the transitional area between the Sudanian and the Ethiopian biogeographical regions (Linder et al. 2012), with different geological and environmental histories (Mayaux et al. 2004). Whereas the former is composed of extensive plateaus of shallow altitude and denser savanna woodlands, the latter represents the northern extreme of the Afromontane region of East Africa, making the transition near the Equator from tropical into more arid climates, across Kenya, dominated by grasslands (Mayaux et al. 2004). A strong hemispheric asymmetry is also verified for large herbivores across the African continent in terms of ecological specialization (Fernández and Vrba 2005), with biome density leading to increased ecological generalization across the Northern hemisphere.

Concordant with previous evaluations of diversity, our genome-wide approach also revealed that the northern cluster exhibits higher levels of genetic diversity than the southern cluster. The observed pattern may reflect the complexity of the evolutionary history across roan's northern distributional range. We hypothesise that an old and larger population faced several vicariant episodes across semi-permeable barriers, for example, vegetation discontinuities, such as the Dahomey Gap (Salzmann and Hoelzmann 2005) or hydrographic systems, such as rivers or lakes (Goudie 2005; Shettima et al. 2018), that acting alone or in combination, occasionally hindered roan's gene flow in this region. Such scenario has been considered for several African taxa, including mammals, birds, and plants (Nicolas et al. 2011; Dobigny et al. 2013; Fuchs and Bowie 2015; Demenou, Doucet and Hardy 2018). On the other hand, genome-wide information did not disclose further sub-structuring in the southern cluster, which contrasts with nuclear diversity assessed using microsatellite loci (Gonçalves et al. this thesis). The faster mutation rate of microsatellites (Li et al. 2002; Väli et al. 2008) and a much larger sampling size may have allowed those authors to capture more recent events of population structuring, that are still difficult to support in the present study. Additionally, a signature of admixture observed in between northern and southern clusters may indicate the presence of a possible contact zone, possibly extending from Central African Republic towards Ethiopia. As previously suggested by Gonçalves et al. (this thesis), this likely reflects interplays of population isolation and secondary contacts

during cyclic climatic and hydrographic changes in the Pleistocene, eventually related to the Nile basin rearrangements.

Sable antelope (*Hippotragus niger*)

Genomic data supports five geographically partitioned clusters, mostly concordant with the species distribution and described subspecies (Ansell 1972; IUCN SSC Antelope Specialist Group 2017b). Population structural congruence across mitochondrial (Pitra et al. 2002, 2006; Rocha 2014; Jansen van Vuuren et al. 2010), microsatellite (Vaz Pinto 2018), and whole-genome data, provides compelling evidence for the recognition of these clusters, across well-defined geographical barriers, namely the EARS across East Africa, and the Zambezi River in the South. The permeability of such geographic barriers is indicated by some levels of admixture detected between individuals from Mozambique, across the Zambezi River, and was previously detected with nuclear data, and associated with unstable physiography (Vaz Pinto 2018). Such strong intraspecific division is most likely related to high spatial heterogeneity across topographically and biogeographically complex regions of East and Southern Africa (Badgley 2010; Linder et al 2012). Checking for signatures of isolation by distance at the genomic level, a significant and positive correlation was found within the Eastern and Southern clusters. This was previously observed for the same groups using a larger sample size, but a much smaller representation of the nuclear genome (Vaz Pinto 2018).

Also in agreement with recent studies (Rocha 2014; Vaz Pinto 2018), our genome-wide approach retrieved higher levels of diversity and divergence for West Tanzanian individuals. The highly divergent mitochondrial lineage that characterizes sable antelopes in this region was hypothesized to be the result of an ancient introgression with an extinct taxon (Pitra et al. 2006; Rocha 2014; Vaz Pinto 2018). Potential imprint of such introgression at the genomic level, however, requires further validation. The present study also reports high levels of diversity for the Zambian cluster, something never achieved using other molecular markers. With a central geographical position across Southern Africa, the Zambian plateau is a geomorphologically stable area, near two major contemporary wetlands, the Zambezi River and the Okavango delta, as well as former paleolakes (Goudie 2005; Burrough and Thomas 2008; Burrough, Thomas and Bailey 2009; Moore, Cotterill and Eckardt 2012). It is possible that such characteristics may have enabled the persistence of a viable population of sable antelopes in this region, facilitating subsequent range expansions. This scenario was also recently reported for the plain zebra, using genomic data (Pedersen et al. 2018). The isolated Angolan population, corresponding to the described subspecies of Giant

sable presents contrasting low levels of genome-wide diversity. This is in line with previous studies (Vaz Pinto 2018) and reflects the population history and current conservation status. After suffering a severe genetic bottleneck during Angolan Civil War, in the 1970s (Pitra et al. 2006), this population was considered as “Critically Endangered” by the Red List of Threatened Species (IUCN SSC Antelope Specialist Group 2008), and it was only saved from the brink of extinction due to a robust management program (Vaz Pinto 2009, 2018). The tendency of loss of genetic diversity after a population bottleneck and consequences therein has been reported recently at the genomic level for other mammals (Abascal et al. 2016; Kardos et al. 2018; Westbury et al. 2018).

4.4.2 Drivers of species diversity and co-occurrence

As sister species, the roan and the sable antelope share many life-history characteristics and habitat adaptations to African savannas. However, several biological and ecological differences may be related to observed differential patterns of distribution and genome-wide diversity. In terms of biology, both species are polygamic, living in breeding herds with a strong social structure. However, roan antelope herds have wider home ranges, and their composition can change daily, whereas sable antelope maternal herds have reduced dispersion over smaller home ranges (Chardonnet and Crosmary 2013; Estes 2013). Dispersion ability can increase the encounter rates between individuals, and for species with multiple reproductive efforts over a lifetime, dispersal can increase the rate of sexual encounters, matting, and reproduction between populational demes (Li and Koko 2018). Contrary to sable antelopes, whose reproductive peak corresponds with the onset of the rainy season, roan antelopes have aseasonal breeding (Chardonnet and Crosmary 2013; Estes 2013; Ogutu, Piepho and Dublin 2015). It is possible that outputted higher levels of genome-wide diversity for the roan antelope, when compared to its congeneric (McLoughlin et al. 2007; Vander Wal, van Beest and Brook 2013), are related to increased opportunities to reproductive events between herds, due to home-range overlap especially mediated by dispersing males, particularly under certain circumstances.

In terms of ecological specialization, the roan antelope has a more generalist nature, with a distribution range that covers several biogeographical biomes (Chardonnet and Crosmary 2013; Linder et al. 2012), whereas the sable antelope is specialized to the miombo biome across Southern Africa (Estes 2013). For large African herbivores, habitat generalization across the Southern hemisphere is positively related to continental area (Fernández and Vrba 2005). Southwards of the vast Zambezian biogeographical region (White 1983; Linder et al. 2012), and with increasing latitude, several smaller biomes

emerge. It is also known that for population density is negatively correlated to body size in mammals, due to a decrease in the habitat carrying capacity (Savage et al. 2004; Bernhardt, O'connor and Sunday 2018). Therefore, in Southern Africa, it is possible that for large herbivores to be able to sustain viable populations, its geographical ranges need to span across several biomes. Spatially heterogeneous environments promote habitat specialization, through disruptive selection (Pontarp et al. 2015). This most likely influences the strong association of sable antelopes to the miombo woodland biome, increasing spatial partitioning, and leading to lower levels of genome-wide diversity, when compared to its congeneric.

Biological and ecological factors inherent to both roan and sable antelope may also facilitate their distributional overlap across East and Southern Africa. Differential patterns of seasonal breeding can promote genetic division, as demonstrated for the giraffe (Thomassen et al. 2013). As a seasonal breeder, regional differences in rainfall availability can disrupt the breeding season between giraffe populations located around the Equator. Region seasonality can also influence dietary niches, and despite being primarily grazers, roan and sable antelopes can also be mixed feeders, depending on seasonal availability (Codron et al. 2007; Chardonnet and Crosmary 2013; Estes 2013). Through DNA metabarcoding, it was assessed that large African herbivores sharing similar diets, can still present marked differences in dietary richness and composition (Kartzinel et al. 2015). This may increase niche opportunities and reduce competition for sympatric species (Stewart et al. 2002; van Beest et al. 2014; Weber and Strauss 2016). Nevertheless, it is possible that currently co-occurring species may have evolved in allopatry, became locally adapted to different habitats, and only after, their distribution became overlapped through range expansion (Estevo et al. 2017). This hypothesis may apply to the roan and sable antelope, considering that the most probable refugial areas for each species are geographically apart (Northern and Southern Africa, respectively), and that African savannas, in general, as well as the miombo woodland biome, in particular, have shifted their distribution throughout both species' evolutionary history (DeMenocal 2004; Lorenzen et al. 2012; Boom et al. 2021).

4.5 References

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J. Y., et al., 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17: 1–19.

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Ansell, W. F. H., 1972. Order Artiodactyla, pp. 15–83 in *Mammals of Africa: an Identification Manual*. Smithsonian Institute Press (J. Meester and H. W. Setzer, Eds.), Washington, DC.

Ayoola, A. O., Zhang, B-L., Meisel, R. P., Nneji, L. M., Shao, Y., Morenikeji, O. B., Adeola, A. C., Ng'ang'a, S. I., et al., 2020. Population genomics reveals incipient speciation, introgression, and adaptation in the African Mona Monkey (*Cercopithecus mona*). *Molecular Biology and Evolution*, msaa248.

Badgley, C., 2010. Tectonics, topography, and mammalian diversity. *Ecography*, 33: 220–231.

Bernhardt, J. R., Sunday, J. M. and O'Connor, M. I., 2018. Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *The American Naturalist*, 192: 687–697.

Bertola, L. D., Jongbloed, H., Van Der Gaag, K. J., de Kniijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., et al., 2016. Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the Lion (*Panthera leo*). *Scientific Reports*, 6: 30807.

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

Black IV, W. C., Baer, C. F., Antolin, M. F. and DuTeau, N. M., 2001. Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, 46: 441–469.

Bolger, A. M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30: 2114–2120.

Bonenfant, C., Gaillard, J-M., Coulson, T., Festa-Bianchet, M., Loison, A., Garel, M., Loe, L. E., Blanchard, P., et al., 2009. Empirical evidence of density-dependence in populations of large herbivores. *Advances in Ecological Research*, 41: 313–357.

Boom, A. F., Migliore, J., Kaymak, E., Meerts, P. and Hardy, O. J., 2021. Plastid introgression and evolution of African miombo woodlands: new insights from the plastome-based phylogeny of *Brachystegia* trees. *Journal of Biogeography*, jbi.14051.

Bowyer, R. T., Bleich, V. C., Stewart, K. M., Whiting, J. C. and Monteith, K. L., 2014. Density dependence in ungulates: a review of causes, and concepts with some clarifications. *California Fish and Game*, 100: 550–572.

Burrough, S. and Thomas, D., 2008. Late Quaternary lake-level fluctuations in the Mababe Depression: Middle Kalahari palaeolakes and the role of Zambezi inflows. *Quaternary Research*, 69: 388–403.

Burrough, S., Thomas, D. and Bailey, R., 2009. Mega-Lake in the Kalahari: a Late Pleistocene record of the Palaeolake Makgadikgadi system. *Quaternary Science Reviews*, 28: 1392–1411.

Chamaillé-Jammes, S., Fritz, H., Valeix, M., Murindagomo, F. and Clobert, J., 2008. Resource variability, aggregation and direct density dependence in an open context: the local regulation of an African elephant population. *Journal of Animal Ecology*, 77: 135–144.

Chardonnet, P. and Crosmary W., 2013. *Hippotragus equinus* Roan Antelope, pp. 548–556 in Mammals of Africa VI. Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., et al., 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, 364: eaav6202.

Codron, D., Codron, J., Lee-Thorp, J. A., Sponheimer, M., De Ruiter, D., Sealy, J., Grant, R. and Fourie, N., 2007. Diets of savanna ungulates from stable carbon isotope composition of faeces. *Journal of Zoology*, 273: 21–29.

Coleman, M. L. and Chisholm, S. W., 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences*, 107: 18634–18639.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, 52: 87–94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763.

DeMenocal, P. B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, 220: 3–24.

Demenou, B. B., Doucet, J. L. and Hardy, O. J., 2018. History of the fragmentation of the African rain forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*. *Heredity*, 120: 547–561.

Demos, T. C., Peterhans, J. C. K., Joseph, T. A., Robinson, J. D., Agwanda, B. and Hickerson, M. J., 2015. Comparative population genomics of African montane forest mammals support population persistence across a climatic gradient and Quaternary climatic cycles. *Plos One*, 10: e0131800.

Diniz-Filho, J. A. F., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., Telles, M. P. D. C., Rangel, T. F. and Bini, L. M., 2013. Mantel test in population genetics. *Genetics and Molecular Biology*, 36: 475–485.

Dobigny, G., Tatard, C., Gauthier, P., Ba, K., Duplantier, J. M., Granjon, L. and Kergoat, G. J., 2013. Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan open habitats Pleistocene history. *Plos One*, 8: e77815.

Ellegren, H., 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29: 51–63.

Estes, R., 2013. *Hippotragus niger* Sable antelope, pp. 556–565 in *Mammals of Africa VI.* Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Estevo, C. A., Nagy-Reis, M. B. and Nichols, J. D., 2017. When habitat matters: habitat preferences can modulate co-occurrence patterns of similar sympatric species. *Plos One*, 12: e0179489.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32: 3047–3048.

Fernández, M. H. and Vrba, E. S., 2005. Rapoport effect and biomic specialization in African mammals: revisiting the climatic variability hypothesis. *Journal of Biogeography*, 32: 903–918.

Fox, E. A., Wright, A. E., Fumagalli, M. and Vieira, F. G., 2019. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, 35: 3855–3856.

Fuchs, J. and Bowie, R. C., 2015. Concordant genetic structure in two species of woodpecker distributed across the primary West African biogeographic barriers. *Molecular Phylogenetics and Evolution*, 88: 64–74.

Fumagalli, M., Vieira, F. G., Korneliussen, T. S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A. and Nielsen, R., 2013. Quantifying population genetic differentiation from Next-Generation Sequencing data. *Genetics*, 195: 979–992.

Goudie, A. S., 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67: 437–456.

Hill, W. G. and Weir, B. S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, 33: 54–78.

IUCN SSC Antelope Specialist Group, 2008. *Hippotragus niger* ssp. *variani*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10169A50188611.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017a. *Hippotragus equinus*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10167A50188287.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017b. *Hippotragus niger*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10170A50188654.en on April 8th, 2020,

Jansen van Vuuren, B., Robinson, T. J., Vaz Pinto, P., Estes, R. and Matthee, C. A., 2010. Western Zambian sable: are they a geographic extension of the giant sable antelope? *South African Journal of Wildlife Research*, 40: 35–42.

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C. and Ellegren, H., 2018. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, 2: 124–131.

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., Rubenstein, D. I., Wang, W. and Pringle, R. M., 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences*, 112: 8019–8024.

Kentie, R., Clegg, S. M., Tuljapurkar, S., Gaillard, J. M. and Coulson, T., 2020. Life-history strategy varies with the strength of competition in a food-limited ungulate population. *Ecology Letters*, 23: 811–820.

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., Grarup, N., Jiang, T., et al., 2011. Estimation of allele frequency and association mapping using Next-Generation Sequencing data. *BMC Bioinformatics*, 12: 231–247.

Kircher, M., Sawyer, S. and Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40: e3–e3.

Koepfli, K.-P., Tamazian, G., Wildt, D., Dobrynin, P., Kim, C., Frandsen, P. B., Godinho, R., Yurchenko, A. A., et al., 2019. Whole genome sequencing and re-sequencing of the Sable antelope (*Hippotragus niger*): a resource for monitoring diversity in *ex-situ* and *in-situ* populations. *G3:Genes, Genomes, Genetics*, 9: 1785–1793.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. and Mayrose, I., 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15: 1179–1191.

Korneliussen, T. S., Albrechtsen, A. and Nielsen, R., 2014. ANGSD: analysis of next

generation sequencing data. *BMC Bioinformatics*, 15: 356–369.

Korneliussen, T. S. and Moltke, I., 2015. NgsRelate: a software tool for estimating pairwise relatedness from Next-Generation Sequencing data. *Bioinformatics*, 31: 4009–4011.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35: 4453–4455.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Segurel, L., Venkat, A., Andolfatto, P. and Przeworski, M., 2012. Revisiting an old riddle: what determines genetic diversity levels within species?. *Plos Biology*, 10: e1001388.

Lefort, V., Desper, R. and Gascuel, O., 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, 32: 2798–2800.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25: 2078–2079.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint*: 1303–3997.

Li, X. Y., Korol, A. B., Fahima, T., Beiles, A. and Nevo, E., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11: 2453–2465.

Li, X. Y. and Kokko, H., 2019. Sex-biased dispersal: a review of the theory. *Biological Reviews*, 94: 721–736.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J. and Rahbek, C., 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39: 1189–1205.

Lorenzen, E. D., Heller, R. and Siegismund, H. R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21: 3656–3670.

Luikart, Gordon, England, P. R., Tallmon, D., Jordan, S. and Taberlet, P., 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4: 981–994.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27: 209–220.

Matthee, C. A. and Robinson, T. J., 1999. Mitochondrial DNA population structure of roan and sable antelope; implications for the translocation and conservation of the species. *Molecular Ecology*, 8: 227–238.

Mayaux, P., Bartholomé, E., Fritz, S. and Belward, A., 2004. A new land-cover map of

Africa for the year 2000. *Journal of Biogeography*, 31: 861–877.

McCullough, D. R., 1999. Density dependence and life-history strategies of ungulates. *Journal of Mammalogy*, 80: 1130–1146.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20: 1297–1303.

McLoughlin, P. D., Gaillard, J. M., Boyce, M. S., Bonenfant, C., Messier, F., Duncan, P., Delorme, D., Van Moorter, B., Saïd, S. and Klein, F., 2007. Lifetime reproductive success and composition of the home range in a large herbivore. *Ecology*, 88: 3192–3201.

Meyer, M. and Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010: pdb.prot5448.

Miller, J. M., Hallager, S., Monfort, S. L., Newby, J., Bishop, K., Tidmus, S. A., Black, P., Houston, B., et al., 2011. Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). *Conservation Genetics*, 12: 423–431.

Moore, A. E., Cotterill, F. and Eckardt, F., 2012. The evolution and ages of Makgadikgadi palaeo-lakes: consilient evidence from Kalahari drainage evolution south-central Africa. *South African Journal of Geology*, 115: 385–413.

Nicolas, V., Missoup, A. D., Denys, C., Kerbis Peterhans, J., Katuala, P., Couloux, A. and Colyn, M., 2011. The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *Journal of Biogeography*, 38: 191–207.

Nielsen, R., Paul, J. S., Albrechtsen, A. and Song, Y. S., 2011. Genotype and SNP calling from Next-Generation Sequencing data. *Nature Reviews Genetics*, 12: 443–451.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. and Wang, J., 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One*, 7: e37558.

Ogutu, J. O., Piepho, H. P. and Dublin, H. T., 2015. Reproductive seasonality in African ungulates in relation to rainfall. *Wildlife Research*, 41: 323–342.

Okonechnikov, K., Conesa, A. and García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32: 292–294.

Oksanen, J. F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., et al., 2019. VEGAN: community ecology package. R package

version 2.5-4. https://CRAN.R-project.org/package=vegan.

Pedersen, C. E. T., Albrechtsen, A., Etter, P. D., Johnson, E. A., Orlando, L., Chikhi, L., Siegismund, H. R and Heller, R., 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology & Evolution*, 2: 491–498.

Pitra, C., Hansen, A. J., Lieckfeldt, D. and Arctander, P., 2002. An exceptional case of historical outbreeding in African sable antelope populations. *Molecular Ecology*, 11: 1197–1208.

Pitra, C., Vaz Pinto, P., O'Keeffe, B. W. J., Willows-Munro, S., Jansen van Vuuren, B. and Robinson, T. J., 2006. DNA-led rediscovery of the giant sable antelope in Angola. *European Journal of Wildlife Research*, 52: 145–152.

Pontarp, M., Ripa, J. and Lundberg, P., 2015. The biogeography of adaptive radiations and the geographic overlap of sister species. *The American Naturalist*, 186: 565–581.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. and Buckler, E. S., 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, 98: 11479–11484.

Rocha, J., 2014. The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences. MsC thesis dissertation, University of Porto, Portugal.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., et al., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515: 261–263.

Salzmann, U. and Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene*, 15: 190–199.

Savage, V. M., Gillooly, J. F., Brown, J. H., West, G. B. and Charnov, E. L., 2004. Effects of body size and temperature on population growth. *The American Naturalist*, 163: 429–441.

Shettima, B., Kyari, A., Aji, M. and Adams, F., 2018. Storm and tide influenced depositional architecture of the Pliocene–Pleistocene Chad Formation, Chad Basin (Bornu Sub–basin) NE Nigeria: a mixed fluvial. *Journal of African Earth Sciences*, 143: 309–320.

Sinding, M. H. S., Gopalakrishan, S., Vieira, F. G., Samaniego Castruita, J. A., Raundrup, K., Heide Jørgensen, M. P., Meldgaard, M., Petersen, B., et al., 2018. Population genomics of grey wolves and wolf-like canids in North America. *Plos Genetics*, 14: e1007745.

Skotte, L., Korneliussen, T. S. and Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195: 693–702.

Slatkin, M., 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47: 264–279.

Smitz, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., Caron, A., Prins, et al., 2013. Pan-African genetic structure in the African Buffalo (*Syncerus caffer*): investigating intraspecific divergence. *Plos One*, 8: e56235.

Stewart, K. M., Bowyer, R. T., Kie, J. G., Cimon, N. J. and Johnson, B. K., 2002. Temporospatial distributions of elk, mule deer, and cattle: resource partitioning and competitive displacement. *Journal of Mammalogy*, 83: 229–244.

Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., Egholm, M., Rothberg, J. M., et al., 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences*, 103:13578–13584.

Swillens, S., Goffard, J. C., Marechal, Y., de Kerchove d'Exaerde, A. and El Housni, H., 2004. Instant evaluation of the absolute initial number of cDNA copies from a single real-time PCR curve. *Nucleic Acids Research*, 32:e56–e56.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123: 585–595.

Thomassen, H. A., Freedman, A. H., Brown, D. M., Buermann, W. and Jacobs, D. K., 2013. Regional differences in seasonal timing of rainfall discriminate between genetically distinct East African giraffe taxa. *Plos One*, 8: e77191.

Väli, Ü., Einarsson, A., Waits, L. and Ellegren, H., 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, 17: 3808–3817.

van Beest, F. M., McLoughlin, P. D., Vander Wal, E. and Brook, R. K., 2014. Density-dependent habitat selection and partitioning between two sympatric ungulates. *Oecologia*, 175: 1155–1165.

Vander Wal, E., van Beest, F. M. and Brook, R. K., 2013. Density-dependent effects on group size are sex-specific in a gregarious ungulate. *Plos One*, 8: e53777.

Vaz Pinto, P., 2009. Giant sable rescue and translocation. *Gnusletter*, 28: 8–10.

Vaz Pinto, P., 2018. Evolutionary history of the critically endangered giant sable antelope (*Hippotragus niger variani*). Insights into its phylogeography, population genetics, demography and conservation. PhD thesis dissertation, University of Porto, Portugal.

Vieira, F. G., Lassalle, F., Korneliussen, T. S. and Fumagalli, M., 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological*

Journal of the Linnean Society, 117: 139–149.

Wang, G., Hobbs, N. T., Twombly, S., Boone, R. B., Illius, A. W., Gordon, I. J. and Gross, J. E., 2009. Density dependence in northern ungulates: interactions with predation and resources. *Population Ecology*, 51: 123–132.

Watterson, G. A., 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7: 256–276.

Weber, M. G. and Strauss, S. Y., 2016. Coexistence in close relatives: beyond competition and reproductive isolation in sister taxa. *Annual Review of Ecology, Evolution, and Systematics*, 47: 359–381.

Westbury, M. V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D. M., Sicks, F., et al., 2018. Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. *Molecular Biology and Evolution*, 35: 1225–1237.

Winter, S., Fennessy, J. and Janke, A., 2018. Limited introgression supports division of giraffe into four species. *Ecology and Evolution*, 8: 10156–10165.

White, F., 1983. *The vegetation of Africa*. United Nations Educational, Scientific and Cultural Organization (Unesco, Ed.), Paris.

Zhang, S., Zhang, Y. and Ma, K., 2016. Latitudinal variation in herbivory: hemispheric asymmetries and the role of climatic drivers. *Journal of Ecology*, 104: 1089–1095.

Zheng, Y., Huang, F., Liang, M., Liu, X. and Yu, S., 2020. The effects of density dependence and habitat preference on species coexistence and relative abundance. *Oecologia*, 194: 673–684.

4.6 Supporting Information

Table S4. 1 Sampling dataset for both roan (*H. equinus*) and sable (*H. niger*) antelope, used in this study, ordered by sampling-code, sample ID, country, and locality. The sampling code used here is included in all the figures throughout the manuscript.

Species	Sampling code	Sample ID	Country	Locality	Gender	Donated by	Original label	Material	Year	N PCR cycles	Median coverage (x)
Roan – (*H. equinus)*	GB1	HeNI217	Guinea Bissau	Kopulan Camp	F	P-C Museum	GUINEA.6	Skin	1911	9	4.3
	GB2	HeNI219	Guinea Bissau	Crocoli Camp	F	P-C Museum	GUINEA.31	Skin	1911	9	4.0
	Gh1	He80	Ghana	Kablima	F	H. R. Siegismund	7324	Muscle	1998	8	5.3
	Gh2	He89	Ghana	Kabanpe	M	H. R. Siegismund	7512	Muscle	1998	9	4.0
	Cm1	He59	Cameroon	Cameroon	M	H. R. Siegismund	4763	Muscle	1997	11	4.5
	Cm2	HeNI229	Cameroon	Bi-Indu	M	P-C Museum	CAMII.117	Skin	1931	5	4.0
	Ch	HeNI222	Chad	Iro Lake	M	P-C Museum	NN.158	Skin	1925	10	4.0
	CAR	HeNI226	CAR	Nana Barya Res.	M	P-C Museum	NN.252	Skin	1925	9	6.0
	Et1	He110	Ethiopia	Gambela Nat. Park	F	H. R. Siegismund	9295	Muscle	—	8	2.7
	Et2	He108	Ethiopia	Akobe septum	M	H. R. Siegismund	9293	Muscle	—	8	12.0
	Ug	He29	Uganda	Kidepo Valley Nat. Park	—	H. R. Siegismund	1234	Muscle	1994	12	3.0
	Ta1	He98	Tanzania	Kigosi Game Res.	M	H. R. Siegismund	8253	Muscle	1997	7	7.2
	Ta2	He95	Tanzania	Maswa Game Res.	M	H. R. Siegismund	8250	Muscle	1997	10	12.7
	Ta3	He102	Tanzania	Ugalla River Game Res.	M	H. R. Siegismund	8634	Muscle	1998	7	5.5
	Ta4	He55	Tanzania	Mlele, Katavi	M	H. R. Siegismund	4127	Muscle	—	10	6.4
	Ta5	He53	Tanzania	Rungwa Game Res.	M	H. R. Siegismund	4123	Muscle	1996	9	6.5
	Za1	HeNI123	Zambia	Chipangali	M	NHM Bulawayo	MAM 22089	Teeth	1963	7	2.1
	Za2	HeNI124	Zambia	Zambia	M	NHM Bulawayo	MAM 22085	Teeth	1963	8	1.5

	SA	HeNI252	South-Africa	Witfonteinrant Mount.	M	Swedish NHM	NRM 581398	Bone	1846	5	4.0
	Na	HeNI205	Namibia	Rundu	M	P-C Museum	SWA.47	Skin	1937	7	4.5
	An1	HeNI207	Angola	Cubango river	F	P-C Museum	SWA.74	Skin	1937	7	14.3
	An2	HeNI211	Angola	Cunene	F	P-C Museum	ANGI.61	Skin	1921	11	4.2
	An3	He26	Angola	Cangandala Nat. Park	M	P. Vaz Pinto	26; Hn150	Muscle	—	6	3.4
Sable – (*H. niger)*	Ta1	Hn185	Tanzania	Luganzo	M	H. R. Siegismund		Muscle		7	12.0
	Ta2	Hn187	Tanzania	Niensi, Katavi	M	H. R. Siegismund		Muscle		6	12.6
	Ta2	Hn275	Tanzania	Niensi, Katavi	M	H. R. Siegismund		Muscle		7	7.1
	Ta2	Hn292	Tanzania	Niensi, Katavi	M	H. R. Siegismund		Muscle		7	7.6
	Ta3	Hn195	Tanzania	Mlele, Katavi	M	H. R. Siegismund		Muscle		6	12.8
	Ta4	Hn307	Tanzania	Lukwati Game Res.	M	H. R. Siegismund		Muscle		6	8.8
	Ta5	Hn295	Tanzania	Rungwa Game Res.	—	H. R. Siegismund		Muscle		6	6.5
	Ta5	Hn299	Tanzania	Rungwa Game Res.	—	H. R. Siegismund		Muscle		5	7.1
	Ke	Hn268	Kenya	Shimba Hills Nat. Res.	F	H. R. Siegismund		Muscle		6	7.4
	Ke	Hn270	Kenya	Shimba Hills Nat. Res.	F	H. R. Siegismund		Muscle		6	7.5
	Ke	Hn276	Kenya	Shimba Hills Nat. Res.	F	H. R. Siegismund		Muscle		10	7.3
	Ke	Hn324	Kenya	Shimba Hills Nat. Res.	M	H. R. Siegismund		Muscle		7	8.5
	Ke	Hn325	Kenya	Shimba Hills Nat. Res.	M	H. R. Siegismund		Muscle		9	12.0
	Ta6	Hn309	Tanzania	Selous Game Res.	—	H. R. Siegismund		Muscle		10	5.4
	Mz1	Hn348	Mozambique	Niassa	M	Jorge Carriço		Skin		8	14.6
	Mz1	Hn286b	Mozambique	Niassa	M	Jorge Carrico		Skin		9	2.3
	Mz2	Hn538	Mozambique	Mahimba	M	B. J. van Vuuren		Skin		8	14.3
	Mz3	Hn340	Mozambique	Gorongosa Nat. Park	M	Marc Stalmans		Muscle		8	6.5

Mz4	MZ040	Mozambique	Marromeu Nat. Res.	M	B. J. van Vuuren		Muscle		9	13.8
Za1	G079	Zambia	Mkushi	M	B. J. van Vuuren		Muscle		8	12.6
Za2	Hn120	Zambia	Zambezi river NW	—	Rubin Els		Muscle		7	15.7
Za3	Hn216	Zambia	Lusaka Nat. Park	M	H. R. Siegismund		Muscle		7	11.4
Za4	Hn315	Zambia	Mumbwa	—	H. R. Siegismund		Muscle		8	6.8
Za5	Hn311	Zambia	Nkala	—	H. R. Siegismund		Muscle		7	6.1
Zm	Hn237	Zimbabwe	Triangle	F	H. R. Siegismund		Muscle		8	12.8
SA	Hn137	South-Africa	Kruger Nat. Park	—	Jeremy Anderson		Muscle		14	6.0
Bo1	Hn202	Botswana	Chobe	M	H. R. Siegismund		Muscle		7	6.7
Bo2	Hn272	Botswana	Linyanti	—	H. R. Siegismund		Muscle		8	8.3
Na	Hn250	Namibia	Mahangu Game Res.	F	H. R. Siegismund		Muscle		7	13.5
Na	Hn255	Namibia	Mahangu Game Res.	—	H. R. Siegismund		Muscle		8	6.2
An1	Hn147	Angola	Luando Natural Res.	F	P. Vaz Pinto		Muscle		7	8.8
An1	Hn163	Angola	Luando Natural Res.	F	P. Vaz Pinto		Muscle		7	7.9
An1	Hn166	Angola	Luando Natural Res.	F	P. Vaz Pinto		Muscle		7	7.3
An1	Hn168	Angola	Luando Natural Res.	F	P. Vaz Pinto		Muscle		9	8.3
An1	Hn169	Angola	Luando Natural Res.	F	P. Vaz Pinto		Muscle		7	6.2
An1	Hn354	Angola	Luando Natural Res.	—	P. Vaz Pinto		Muscle		5	13.2
An1	Hn367	Angola	Luando Natural Res.	—	P. Vaz Pinto		Muscle		6	8.1
An2	PA20	Angola	Cangandala Nat. Park	F	P. Vaz Pinto		Muscle		6	6.1

Sample: code He and Hn – contemporary for roan and sable, respectively; HeNI – historical roan samples

Country: CAR – Central African Republic

Location: Nat. Park – National Park; Res. – Reserve; Mount. – Mountain; Nat. Res. – National Reserve

Gender: M – male; F – female;

Donated by: P-C Museum – Powell-Cotton Museum, Kent, UK; NHM Bulawayo – Natural History Museum Bulawayo, Zimbabwe; Swedish NHM – Swedish Natural History Museum, Stockholm, Sweden

Table S4. 2 Mantel's test for isolation-by-distance within cluster for both roan (*H. equinus*) and sable antelope (*H. niger*). Test using the Pearson correlation model and 10,000 permutations.

Roan (*H. equinus*)

	Mantel r	p-value
Northern	0.434	0.0179*****
Southern	0.422	0.0009*****

Sable (*H. niger*)

	Mantel r	p-value
West Tanzanian	-0.163	0.7072
Eastern	0.867	0.0029*****
Zambian	0.136	0.3416
Southern	0.537	0.0029*****
Angolan	0.597	0.1068

***** significant values for p-value < 0.05.

Table S4. 3 Hardy-Weinberg equilibrium (HWE) test within cluster for both roan (*H. equinus*) and sable antelope (*H. niger*), showing the ratio of site allele frequencies with and without the assumption of HWE.

Roan (*H. equinus*)

	with HWE	without HWE
Northern	0.1830	0.1934
Southern	0.1609	0.1664

Sable (*H. niger*)

	with HWE	without HWE
West Tanzanian	0.2226	0.2245
Eastern	0.2065	0.2093
Zambian	0.2464	0.2488
Southern	0.2218	0.2241
Angolan	0.2527	0.2549

Table S4. 4 Tajima's D (Tajima 1989) values within population for both roan (*H. equinus*) and sable antelope (*H. roan*), averaged across sliding-windows of 100,000 bp, spaced every 25, 000 bp.

Roan (*H. equinus*)

	mean Tajima's D
Northern	-1.14×10^{-6}
Southern	-1.08×10^{-6}

Sable (*H. niger*)

	mean Tajima's D
West Tanzanian	1.65×10^{-7}
Eastern	-4.50×10^{-9}
Zambian	1.59×10^{-7}
Southern	2.15×10^{-7}
Angolan	7.15×10^{-7}

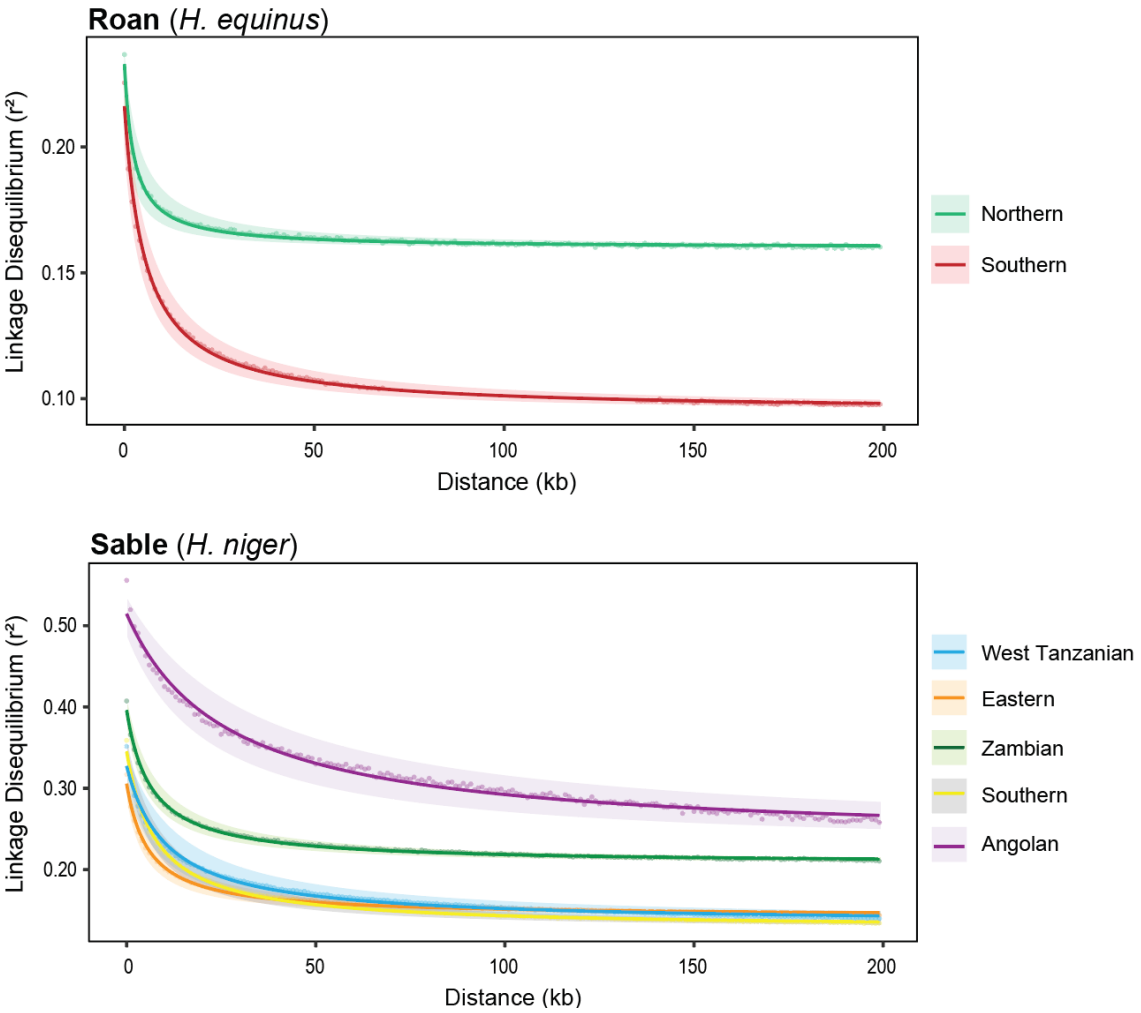


Figure S4. 1 Results of linkage disequilibrium (r^2) decay fitting across genetic distance, for the two clusters of roan antelope (top) and the five clusters of sable antelope (bottom), represented by coloured lines according to legend. Dataset using all individuals within each population, and best fitted curves were based on 1,000 bp bins (points).

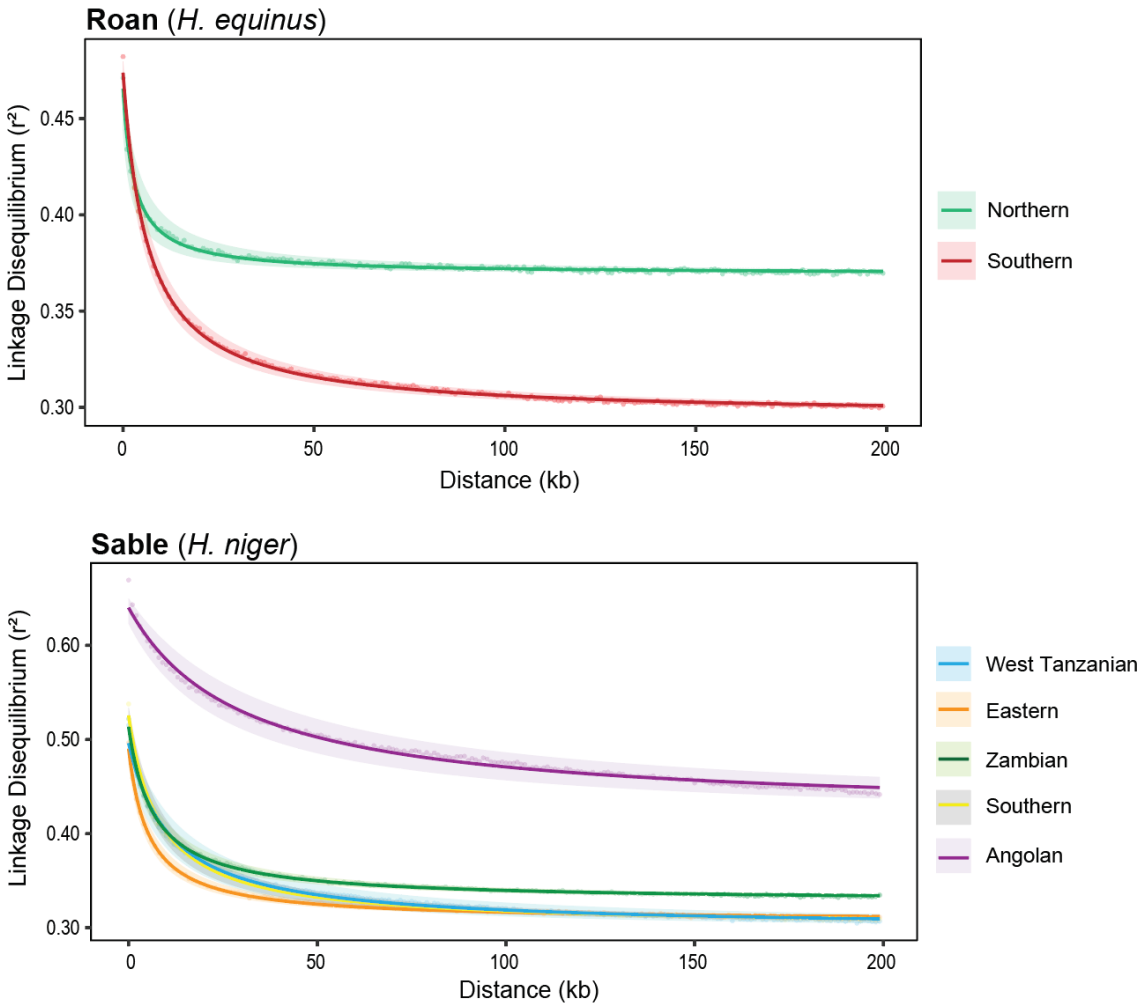


Figure S4. 2 Results of linkage disequilibrium (r^2) decay fitting across genetic distance, for the two clusters of roan antelope (top) and the five clusters of sable antelope (bottom), represented by coloured lines according to legend. Dataset using only the 3 individuals within each population with the highest coverage, and best fitted curves were based on 1,000 bp bins (points).

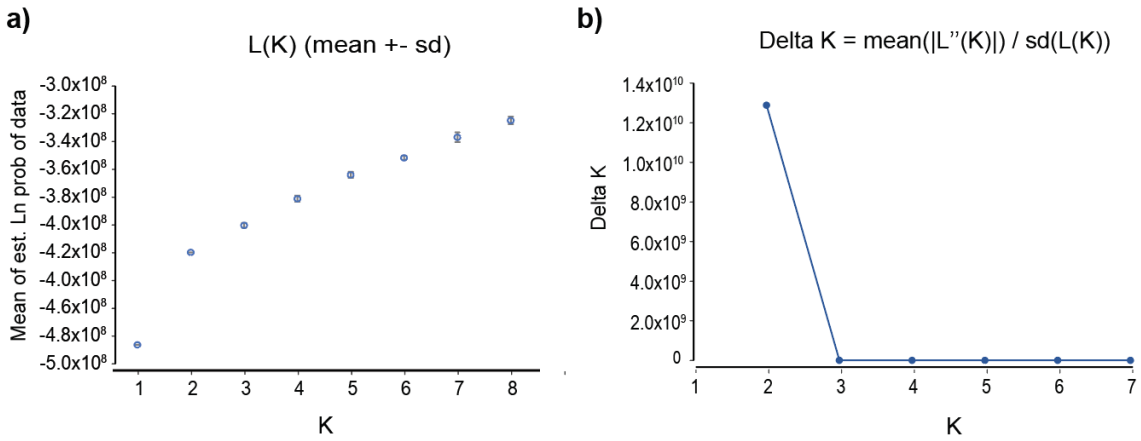


Figure S4. 3 Results from NGSadmix run from K = 1 to K = 8 for roan antelope (*H. equinus*). **a)** posterior likelihood [L(K)] values from 10 independent run per K; **b)** delta-K based on the range of change of log probability for the same analysis.

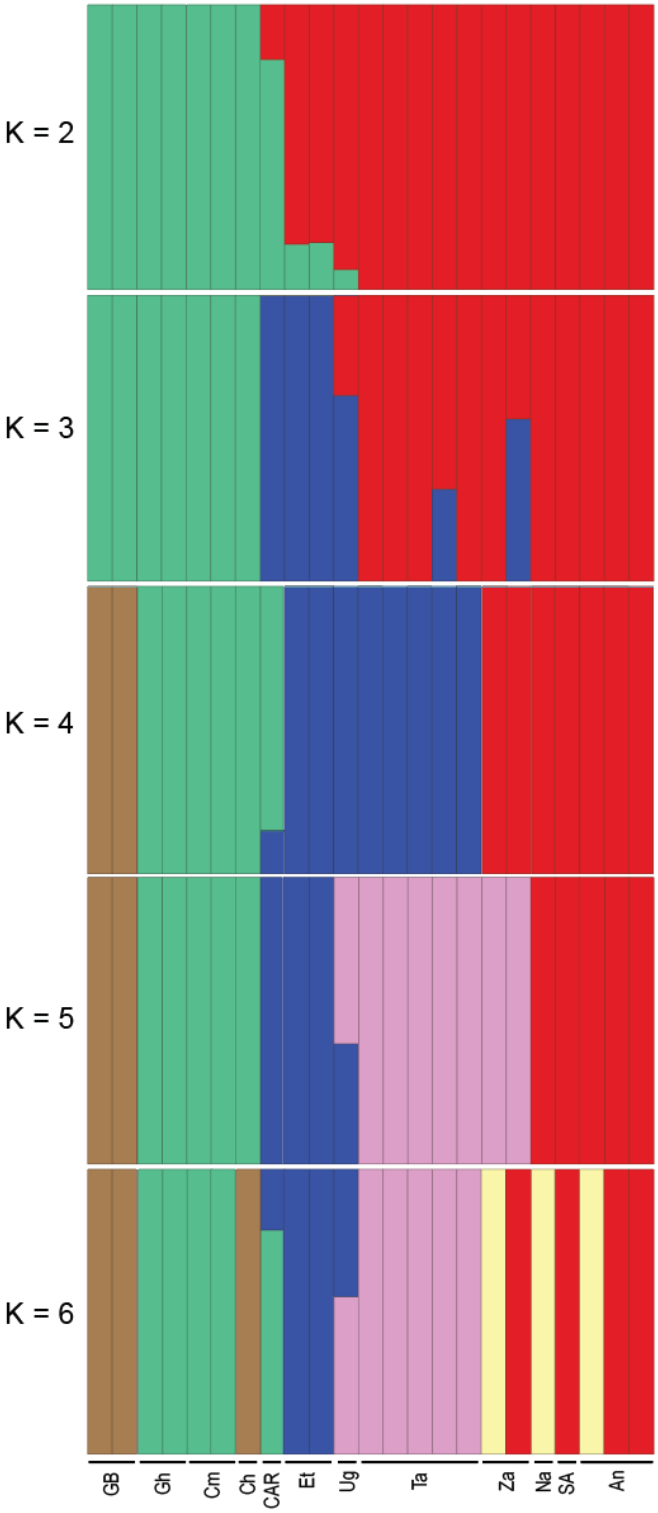


Figure S4. 4 Estimated individual ancestry proportions after NGSAdmix run, from K = 2 to K = 6, for roan antelope (*H. equinus*). Individuals are represented by a vertical bar and each colour represents a different cluster. Individuals ordered according to sampling code, as in Table S4.1, Figure 4.1, Figure 4.3 and Figure 4.5.

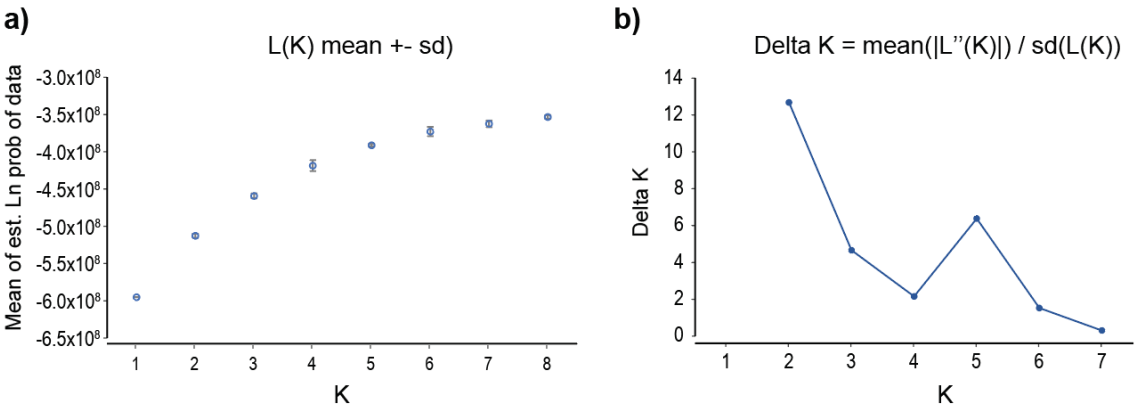


Figure S4. 5 Results from NGSadmix run from K = 1 to K = 8 for sable antelope (*H. niger*). **a)** posterior likelihood [L(K)] values from 10 independent run per K; **b)** delta-K based on the range of change of log probability for the same analysis.

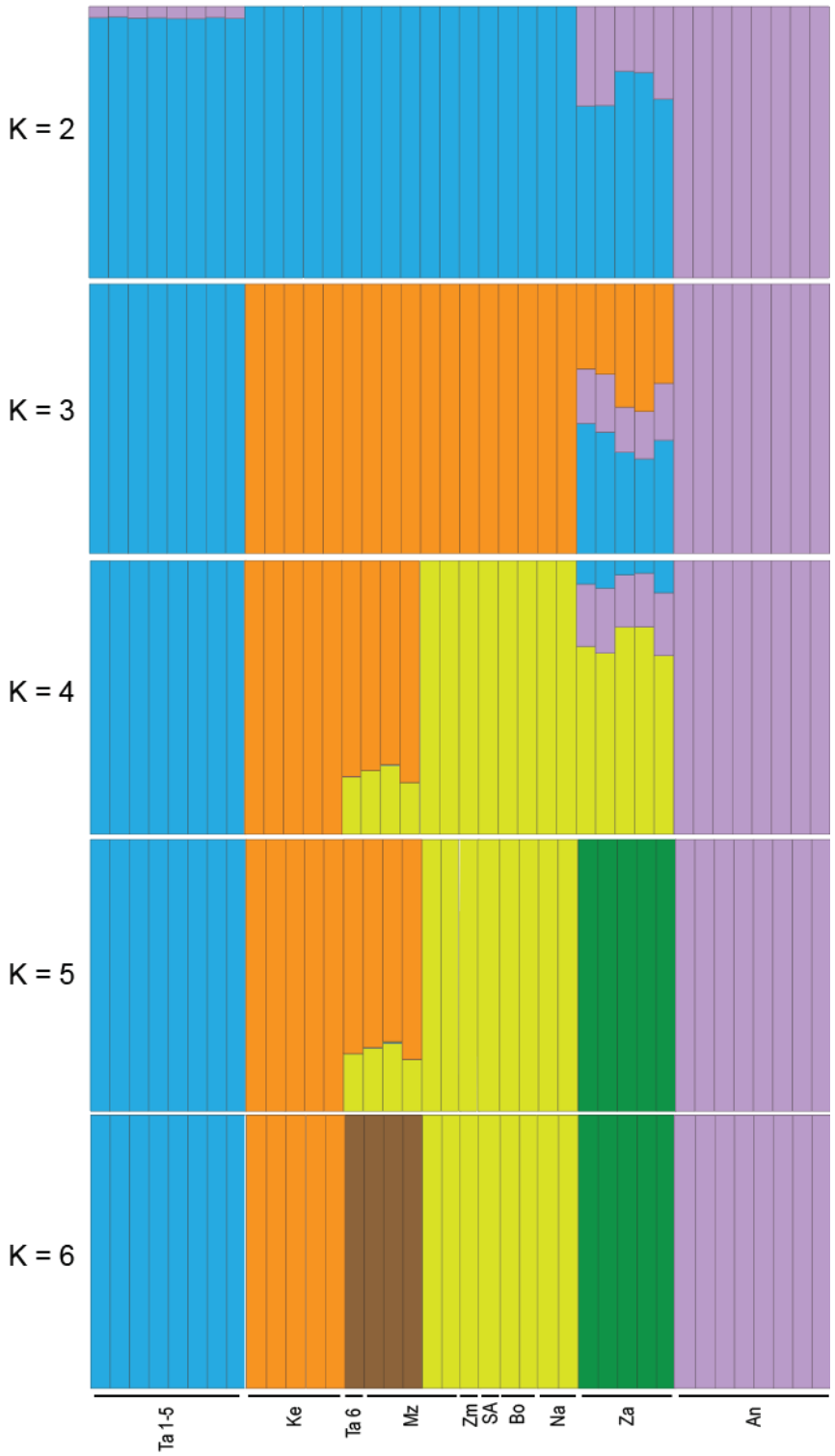


Figure S4. 6 Estimated individual ancestry proportions after NGSAdmix run, from K = 2 to K = 6, for sable antelope (*H. niger*). Individuals are represented by a vertical bar and each colour represents a different cluster. Individuals ordered according to sampling code, as in Table S4.1, Figure 4.2, Figure 4.4 and Figure 4.6.

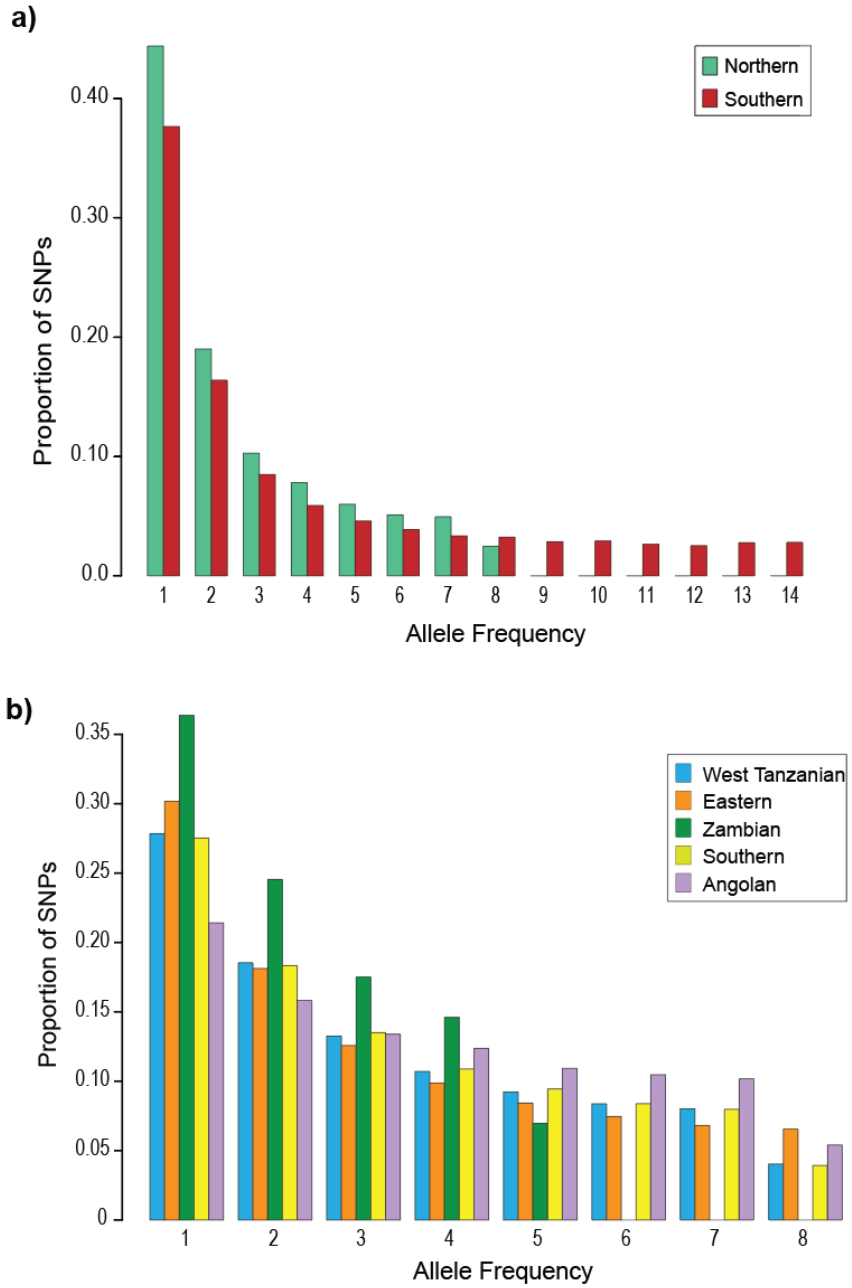


Figure S4. 7 Nucleotide diversity from populational site-frequency spectra, measured as the proportion of SNPs per allele frequency, across sliding windows of 100,000 bp, spaced every 25,000 bp. **a)** for the two clusters of roan antelope (*H. equinus*); **b)** for the five clusters of sable antelope (*H. niger*). Colours according to figure legend.

CHAPTER 5

General Discussion

5.1 Overview of the thesis

The main objective of this thesis was to assess the evolutionary history of extant species of *Hippotragus* genus, the roan (*H. equinus*) and the sable antelope (*H. niger*), in light of extrinsic factors that have been shaping sub-Saharan Africa throughout geological times. Through a comparative analysis, using previously published scientific work, we first reviewed the genetic discontinuities found for fauna and flora taxa across this region (Chapter 2). Congruent patterns across several taxonomic groups, including antelope species, such as those within the *Hippotragus* genus, revealed common evolutionary trajectories in response to African's geomorphologic and climatic history.

Remaining findings were focused on the roan antelope (*H. equinus*), using genetic and genomic approaches towards spatial analysis of genetic partitioning, divergence estimates, and demographic analysis, as well as whole-genome assembly of the first reference genome for this species, and intraspecific genomic assessment of diversity (Chapter 3). Using re-sequencing data for both roan and sable antelope (*H. niger*), we also performed a comparative genomic analysis of the genus, relating differential patterns with both intrinsic and extrinsic factors (Chapter 4).

5.1.1 Sub-Saharan African phylogeography

The African savanna constitutes one of the World's major biomes, covering about half of the continent across sub-Sahara, encompassing grassland, scrubland, and woodland habitats that harbour the largest remaining global megafauna (Fjeldså et al. 2004; Mayaux et al. 2004; Sankaran et al. 2005; Ratnam et al. 2011). It is, therefore, one of the richest areas in the World in terms of biodiversity, particularly for ungulate species (Du Toit and Cumming 1999). Through comparative phylogeographic studies it is possible to uncover common patterns of biodiversity for co-distributed taxa within a particular biome, since they experience similar mechanisms of diversification and evolutionary drivers (Bermingham and Moritz 1998; Arbogast and Kenagy 2008; Avise 2009; Hickerson et al. 2010). Biodiversity patterns across African savannas have been highly correlated with historical climatic and geomorphological changes that occurred during the Pleistocene period (DeMenocal 2004; Hewitt 2004; Dupont 2011; Lorenzen, Heller and Siegismund 2012). After the rapid evolution of savanna-adapted taxa during Pliocene favourable climatic period (Vrba 1996; Bobe 2001; Bibi and Kiessling 2015), African climates became increasingly variable throughout the Pleistocene (DeMenocal 1995; Trauth, Larrasoaña and Mudelsee 2009), with cyclic variation of savanna expansion during interpluvials (cold and drier climate), and savanna recession during

pluvials (warm and wetter climate) (Dupont 2011). At the same time, sub-Saharan Africa witnessed drastic geomorphological changes, with tectonically active regions and drainage topography evolution (Chorowicz 2005; Goudie 2005).

Comparative phylogeographic studies on savanna-adapted taxa, particularly mammal species, present spatially congruent phylogenetic patterns across West, East, and South of Africa (Hewitt 2004; Lorenzen et al. 2012), suggested as main Pleistocene pluvial refugia, that were able to sustain suitable habitat and maintain genetic diversity. The evolutionary history of these taxa was, therefore, most likely driven by speciation within, as well as dispersal events from such refugia, leading to the currently observed biodiversity across the savanna biome. For some taxa, it is suggested the existence of a fourth refugium, across Southwest Africa (Lorenzen et al. 2012).

Based on previous work, we wanted to identify if these phylogenetic patterns would also be congruent for other taxonomic groups, other than mammals, and to uncover other relevant patterns, rarely recognized. For this purpose, we reviewed more than 170 research articles to perform a comparative analysis on the genetic discontinuities for sub-Saharan taxa (Chapter 2). We were able to identify a total of 19 patterns of genetic discontinuity for this region, most of them consistent among several taxonomic groups of fauna (mammals, birds, reptiles, amphibians, and fish) and flora. Our analysis clearly demarked the Tropical Forest belt and the EARS, which were previously used to define the main three African refugial regions to the West, East, and Southern Africa, and divergence time estimates where mainly linked to the Pleistocene period, also in agreement with previous notions (Hewitt 2004; Lorenzen et al. 2012). This most likely reflects common evolutionary trajectories across taxonomic groups, dictated by the position of main geographical barriers (Figure 5.1) and climatic history. In general terms, the Tropical Forest belt, across Central Africa, constitutes a natural barrier for savanna-adapted species, and together with the EARS, demarks a deep genetic differentiation between populations across West Africa, from those to the East and Southern Africa (Chapter 2). The Zambezi River is largely associated with a division between eastern and southern populations, namely for mammal species, of which the *Hippotragus* is included. Remaining 16 patterns were related to other geographic barriers limiting connectivity across each main region, to regional heterogeneity of the landscape, at a finer scale, and to ecological constraints for each taxon (Chapter 2). The lack of patterns of genetic discontinuity was highly associated with high dispersal ability, recent population expansion, or even ongoing hybridization within contact zones (Chapter 2).

Ultimately, the identification of such patterns across savanna-adapted taxa, allows us to better understand the evolutionary mechanisms that have been shaping the biodiversity of this important biome. When applied towards our model species, the roan

and the sable antelope (*Hippotragus* spp.), we were able to detect such congruent patterns, and relate them with common evolutionary histories that resulted from similar biological characteristics and ecological adaptations (next section).

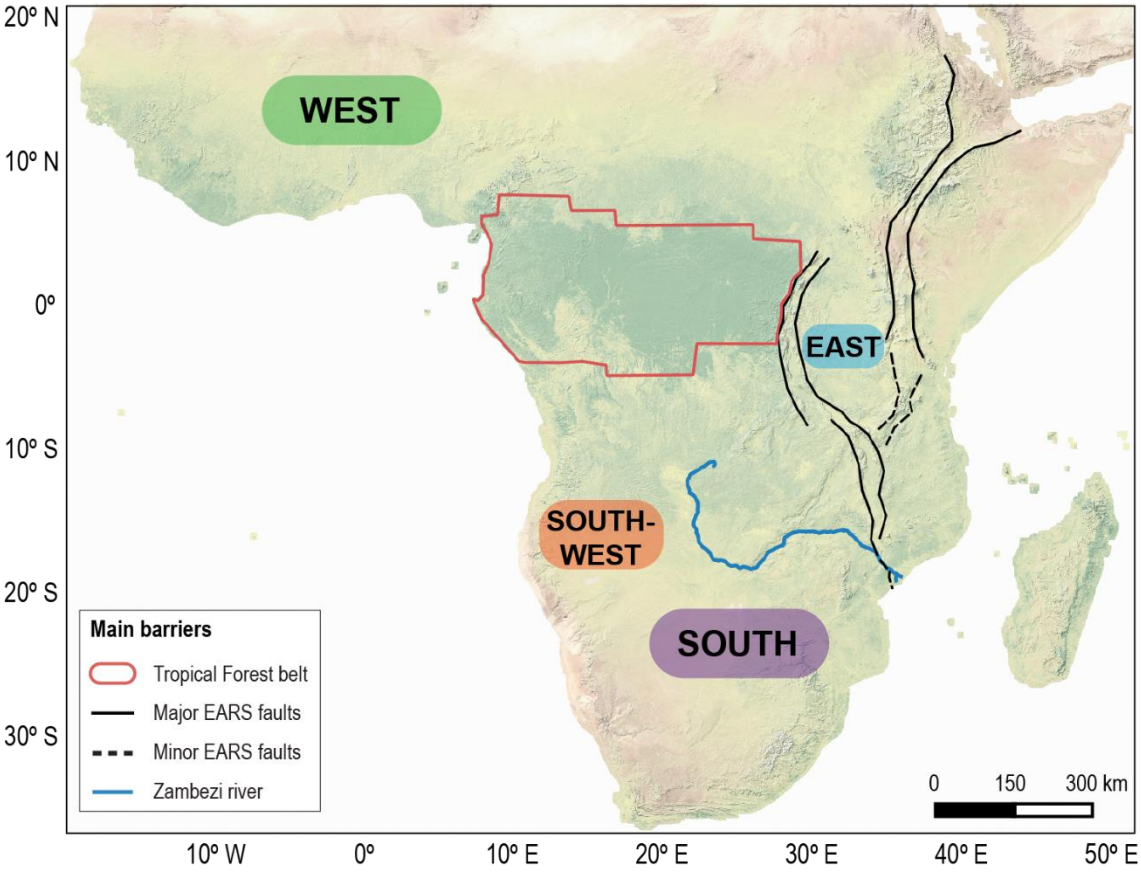


Figure 5. 1 African phylogeographic features among several taxa, namely *Hippotragus* genus, adapted from Hewitt (2004), Lorenzen et al. (2012), as well as analysis performed in this thesis (Chapter 2). Most likely location of main Pleistocene pluvial refuge areas: West (in green); South (in purple); South-West (in red); as well as East (in blue) acting as contact zone between the former. Also represented the main geographic barriers between refugial areas, relevant among several African taxa, but in particular for *Hippotragus* genus: the Tropical Forest belt (red polygon); the East African Rift System – EARS, adapted from Chorowicz (2005) (black curves), with the eastern and western branches represented as major faults, in solid curves, and minor faults in dotted curves, coincident with the Eastern Arc Mountains – EAM; and the Zambezi River (blue curves). Land cover is represented from forests (in green) to mountains (in brown), with shaded relief, depicting higher altitude.

5.1.2 Evolutionary history of *Hippotragus* genus

Hippotragus is an African endemic genus, comprising the extinct bluebuck (*H. leucophaeus*) and two extant species of savanna-dweller antelopes: the roan (*H. equinus*) and the sable (*H. niger*). They share many life-history characteristics (Chardonnet and Crosmary 2013; Estes 2013), being found in sympatry throughout most of their distribution range across sub-Saharan mesic woodland savannas (IUCN SSC Antelope Specialist Group 2017a, 2017b). By mitochondrial data, both species have

diverged around 5.8 Mya (Bibi 2013), presenting different distributional ranges and habitat specialization (Chardonnet and Crosmary 2013; Estes 2013), however with an evolutionary history that is highly associated with the Pleistocene climatic cycles (Lorenzen et al. 2012), constituting, therefore, an optimal model to study African phylogeography.

Both species have been the focus of several studies assessing their levels of population diversity, structure, and differentiation at the intraspecific level, while assessing the validity of proposed subspecies by Ansell, primarily based on geographical distribution, but also on phenotypic characters, such as coat colouration, facial marks, and body size (see Ansell 1972, and references therein).

Roan antelope' pan-African distribution can be divided into six subspecies: *H. e. koba* in the northwest; *H. e. charicus* in North-central Africa; *H. e. bakeri* in the northeast; *H. e. langheldi* to the east; *H. e. cottoni* in South-central Africa; and *H. e. equinus* in the south (Ansell 1972) (Figure 5.2a). Previous studies on genetic data challenged the validity of these subspecies, by proposing the existence of two ESUs, separating the northwest (*H. e. koba*) from remaining specie's distribution (Alpers et al. 2004). In this thesis, through the analyses of powerful genetic markers, both at the genetic and genomic level, together with a comprehensive dataset, we were able to provide better insights on the spatial genetic structure of this species (Figure 5.2).

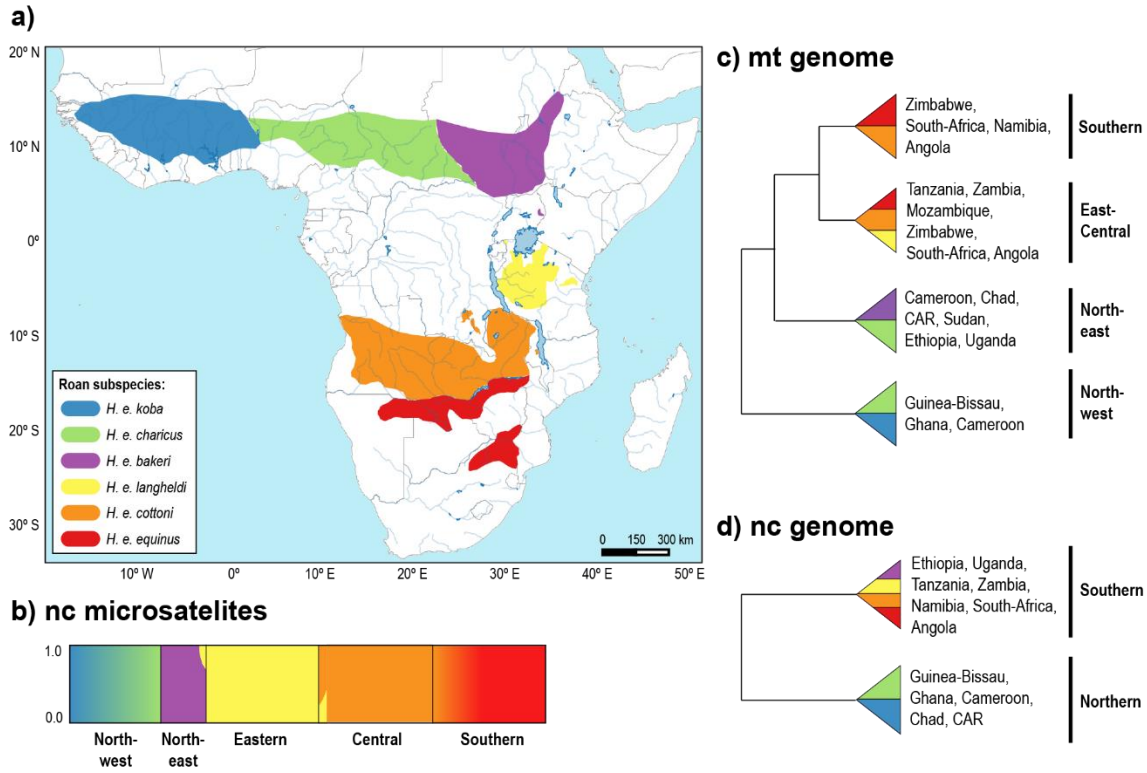


Figure 5. 2 The **r**oan antelope intraspecific analysis. **a)** subspecies geographical distribution, according to Ansell (1972) and current species distribution (IUCN SSC Antelope Specialist Group 2017a); **b)** clustering analysis showing the

assignment proportions of each individual using nuclear microsatellite loci, as analysed in this thesis (Chapter 3 – Study I); **c)** schematic cladogram showing intraspecific phylogenetic relationships using whole-mitochondrial data, as analysed in this thesis (Chapter 3 – Study I); **d)** schematic cladogram showing intraspecific phylogenetic relationships, using whole-genome data, as analysed in this thesis (Chapter 4). mtDNA lineages and nuclear groups identification following the ones from the authors. Each colour corresponds to a subspecies, according to the legend.

Clustering analysis on nuclear microsatellite data (Chapter 3 – Study I), as well as whole-genomic data (Chapter 3 – Study II; Chapter 4), retrieved a genetic partitioning that was mostly consistent with previous studies on the roan antelope (Matthee and Robinson 1999; Alpers et al. 2004), separating West Africa (*H. e. koba* and *H. e. charicus*) from remaining specie's distribution (Figure 5.2d). This pattern is highly associated with the position of the Tropical Forest belt, across Central Africa, and is found across many savanna species with similar distribution ranges to roan, such as ungulates (Lorenzen et al. 2012; Smitz et al. 2013; Winter, Fennessy and Janke 2018), other mammals (Bertola et al. 2016), and even birds (Miller et al. 2011). This division is coincident with the transition between the Northern and the Southern hemispheres, as well as different biogeographical areas (Linder et al. 2012), presenting dissimilar geological histories (Mayaux et al. 2004). From a hierarchical clustering analysis on nuclear microsatellite data (Chapter 3 – Study I), we were able to further partition the northeaster, eastern, and southern distribution of the roan antelope into four clusters, mostly congruent with the described subspecies (Figure 5.2b): north-east (*H. e. bakeri*); eastern (*H. e. langheldi*); central (*H. e. cottoni*); and southern (*H. e. equinus*). The only exception was for the Angolan individuals, genetically similar to those of the nominal *equinus* subspecies, rather than to *H. e. cottoni*. Whole-mitochondrial data (Chapter 3 – Study I) retrieved a similar partition (Figure 5.2c), however with an introgression between individuals to the northwest and northeast, as well as a single mtDNA lineage characterizing individuals in the east and south-central distribution (Figure 5.2c – east-central lineage), with some level of introgression with a fourth southern lineage. This pattern of genetic partitioning, together with mtDNA divergence estimates, can be associated with several events during the Pleistocene period, namely the continued activity of the EARS (Chorowicz 2005) and climatic cycles affecting African vegetational distribution and water systems (Goudie 2005; Dupont 2011). The division between the northwest and northeast, for example, is coincident with the connection between the White and Blue Nile rivers, across central Sudan, during the late-Pleistocene (Goudie 2005; Shettima et al. 2018); the division between northeast and eastern clusters may be associated with the formation of Lake Victoria, in Uganda, and the drainage reversal of the associated Lake Kyoga, around the same period (Goudie 2005); and the differentiation between the central and southern clusters, can be related to the extreme

hydrographic changes that took place during the mid-Pleistocene in Southern Africa (Walford, White and Sydow 2005), including the connection of the Upper and Middle Zambezi rivers.

According to divergence estimates, as well as levels of genetic diversity and differentiation, West Africa acted as a population refugia for roan antelope during the Pleistocene, with range expansions to the East and Southern Africa during favourable periods, across semi-permeable geographical barriers. A similar scenario has also been proposed for other ungulates including the hartebeest, buffalo, and giraffe (Flagstad et al. 2001; Brown et al. 2007; Smitz et al. 2013). Altogether, the results obtained in the thesis, both at the genetic and genomic level, indicate a complex evolutionary history for the roan antelope, with repeated population isolation and subsequent colonisations of new areas through range expansions, during geomorphological transformation, as well as cyclic climatic and hydrographic changes, that occurred in Africa during the Pleistocene period. Based on demographic analysis at the nuclear level (microsatellite data), during the Holocene, roan antelopes started to experience a general trend of population decrease, likely associated with extreme droughts, as well as increased human-pressure at the beginning of the Neolithic. The same trend was detected for other water-dependant ungulate species, such as the African buffalo (Heller et al. 2008; Heller, Brüniche-Olsen and Siegismund 2012).

According to Ansell (1972) and recent data (Vaz Pinto 2018), the distribution of the sable antelope across East and Southern Africa can be divided into five entities (Figure 5.3a). Population structural congruence across mitochondrial (Pitra et al. 2002, 2006; Rocha 2014; Jansen van Vuuren et al. 2010), microsatellite (Vaz Pinto 2018), and whole-genome data (Chapter 4), provides compelling evidence for the recognition of these five entities, mostly consonant with the described subspecies: i) individuals across western Tanzania; ii) *H. n. roosevelti*, across Kenya, south-eastern Tanzania, and northern and central Mozambique; iii) *H. n. kirkii*, in Zambia; iv) *H. n. niger*, across southern Mozambique, Zimbabwe, South-Africa, Botswana, and Namibia; and v) *H. n. variani*, in Angola. This partition, as for the roan antelope, can be highly associated with the continued activity of the EARS during the Pleistocene period (Chorowicz 2005), namely the position of the EAM across the eastern branch, separating the highly divergent West Tanzanian population from remaining ones (Figure 5.3). Across Southern Africa, the hydrographic changes of the Zambezi drainage are related to the division of the southern (*H. n. niger*) population, and the isolation of the Angolan population (*H. n. variani*) is most likely related to habitat heterogeneity at the regional scale (Vaz Pinto 2018).

Even though there seems to be a congruence between proposed subspecies and genetic data for both species, the phenotypic characters used to assist on subspecific

description (see Ansell 1972 and references therein) have never been explicitly quantified for either species. The assemblies of both species' reference genomes (Koepfli et al. 2019; Chapter 3 – Study II) enable a future search for signatures of selection, through direct genomic scans across different populations, or more sophisticated genome-wide association studies (GWAS) of quantitative traits, possibly with helpful outcomes for the understanding of the the molecular drivers that support phenotypic variation (Bouwman et al. 2018; Orteu and Jiggins 2020).

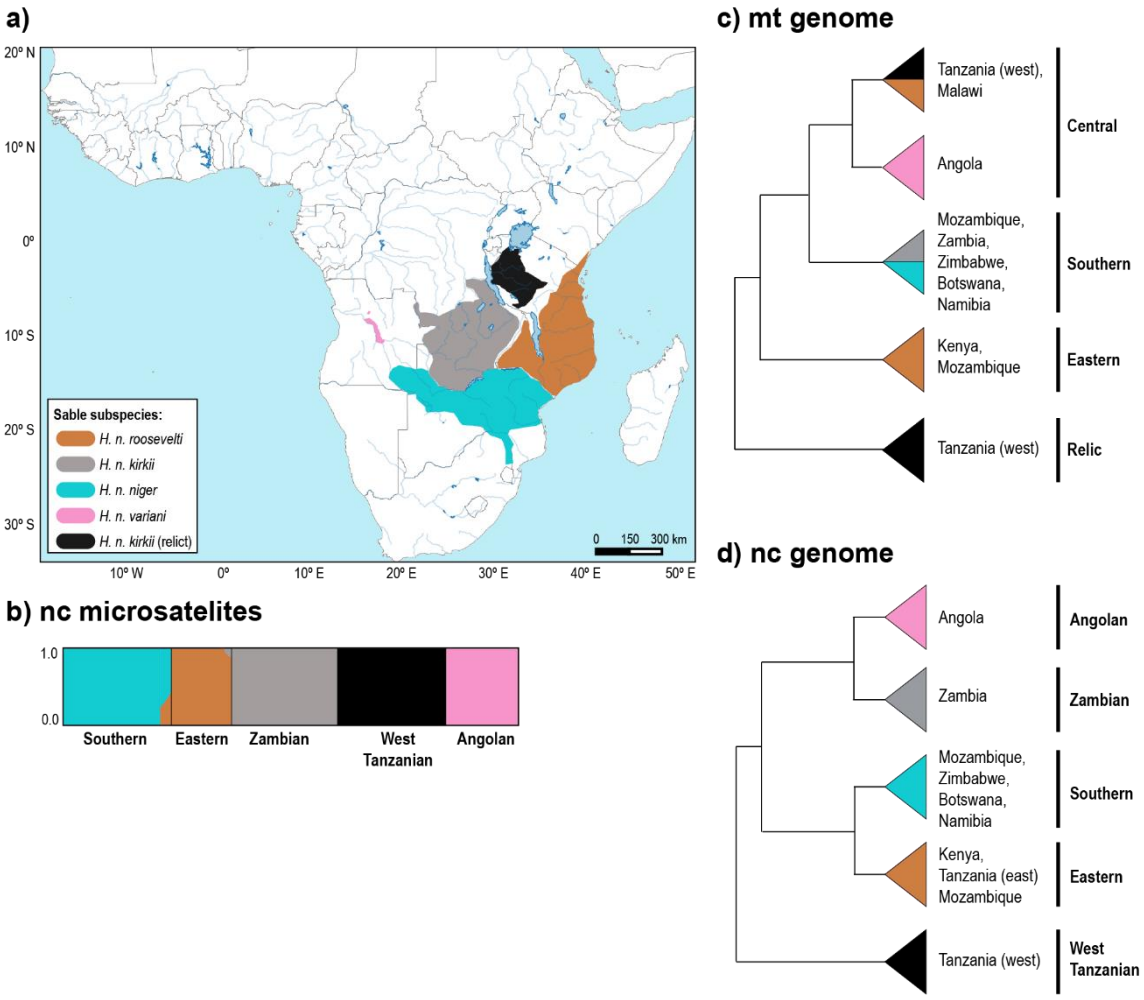


Figure 5. 3 The sable antelope intraspecific analysis. **a)** subspecies geographical distribution, according to Ansell (1972) and current species distribution, adapted from IUCN SSC Antelope Specialist Group (2008, 2017b); **b)** clustering analysis showing the assignment proportions of each individual using nuclear microsatellite loci, adapted from Vaz Pinto (2018); **c)** schematic cladogram showing intraspecific phylogenetic relationships using whole-mitochondrial data, adapted from Rocha (2014); **d)** schematic cladogram showing intraspecific phylogenetic relationships, using whole-genome data, as analysed in this thesis (Chapter 4). mtDNA lineages and nuclear groups identification following the ones from the authors. Each colour corresponds to a subspecies, according to the legend.

By performing a comparative genomic analysis between both species (Chapter 4), we were able to relate differential levels of genetic partitioning and diversity with species-specific features, including behavioural characteristics (Chardonnet and Crosmary 2013;

Estes 2013), such as breeding seasonality and home size range (Ogutu, Piepho and Dublin 2015; Thomassen et al. 2013), as well as, ecological specialization (Fernández and Vrba 2005), influencing range distribution, dietary niches, and facilitating co-distribution (van Beest et al. 2014; Kartzinel et al. 2015; Estevo, Nagy-Reis and Nichols 2017). The evolutionary history of the *Hippotragus* genus enables us to understand how sister species, adapted to the savanna biome, and subjected to the same geological and climatic events, followed differential evolutionary trajectories, and developed several biological and ecological characteristics that diminish their competition when in sympatry. Such inferences can be extrapolated towards other African species with similar characteristics.

5.2 Conservation and management implications

Even though the focus of this thesis was on the evolutionary history of the roan and sable antelopes, results on spatial genetic structure and summary statistics have direct implication towards conservation and management plans. Optimal conservation plans rely on defining the number of distinct spatial units, inferred from the spatial genetic structure within distinctive species, subspecies, or populations (Moritz 1994, 1999). According to the IUCN Red List of Threatened Species, both roan and sable antelope are considered as “Least Concern” (IUCN SSC Antelope Specialist Group 2017a, 2017b), despite decreasing population tendencies over the past few decades and the fact that most wild populations are only sustained within protected areas (East and IUCN SSC Antelope Specialist Group 1999). At the subspecific level, uncertain genetic validity, and vague distribution limits, generally leads to unrecognized units for conservation purposes. Even though the roan and sable antelope have six and four described subspecies, respectively (Ansell 1972), only the Angolan Giant sable (*H. n. vairiani*) is recognized by the IUCN, currently listed as “Critically Endangered” (IUCN SSC Antelope Specialist Group 2008). After suffering a population bottleneck during the Angolan Civil War (1975-2002), this subspecies was only saved from the brink of extinction due to a robust management program (Vaz Pinto 2009). The genetic consequences of this bottleneck were also confirmed in this thesis, at the genome-wide level, with this population showing reduced levels of diversity (Chapter 4). Remaining groups (as described in section 5.1.2) show genomic signals of population stability, and therefore, no relevant concern from a conservation perspective. However, since these groups present a strong geographic partitioning, they should be managed as different units, with retrieved phylogenetic relationships (Chapter 4 – Figure 4.6) being used to assist on management within, as well as translocation decisions between such units.

For the roan antelope, lack of conservation and management plans are of greater concern, given that some populations across its eastern and southern distribution are considered rare and even virtually extinct (East and IUCN SSC Antelope Specialist Group 1999; Chardonnet and Crosmary 2013). Results retrieved in this thesis, namely at the genomic level (Chapter 3 – Study II and Chapter 4), are in agreement with the previous proposal of two ESUs separating the northwest from remaining species' distribution (Alpers et al. 2004). Nevertheless, analyses at the mitochondrial level and on polymorphic microsatellite data (Chapter 3 – Study I), were able to capture further partitioning within the second ESU, separating the north-eastern, eastern, central, and southern roan distribution. Divergence time estimates, together with diversity statistics, indicate a reduction from north-eastern towards southern populations, due to repeated range expansions from refugia. Such an evolutionary scenario, when taken together with reports on population decrease over the past few decades, due to anthropogenic threats (IUCN SSC Antelope Specialist Group 2017a), can drive these populations across East and Southern Africa beyond the point where they cannot respond to change. Expanded surveys across this region are needed to fully access these concerns, however, these regional collapses are already imprinted in the generalized lower genetic diversity in contemporary versus historic samples for the southern roan population (Chapter 3 – Study I). Across this region, as an economically-important species, roan antelope numbers are increasing in private land, where individuals are intensively managed under semi-captive breeding conditions (or game ranching) (Bothma and van Royen 2005; Piltz, Sorensen and Ferrie 2016). Despite the economic advantage, the conservation cost of game ranching can become overwhelming, and better regulation is necessary (Cousins, Sadler and Evans 2010; Pitman et al. 2017). In the past, some of these captive breeding practices included the translocation of roan antelopes across large geographic distances, with subsequent intraspecific hybridization on game farms, as recently reported by van Wyk et al. (2019). This situation creates an entirely new set of considerations, requiring bold and critical ideas around this species' conservation. Genomic resources provided in this thesis (Chapter 3 – Study II) and retrieved phylogenetic relationships (Chapter 4 – Figure 4.5) may prove highly valuable on future decisions, towards sustainability (Wildt et al. 2019).

5.2.1 A case-study on South-African roan antelopes

Reports on increasing numbers of roan antelopes under semi-captive breeding (or game ranching) in South-Africa, contrast highly with the situation across this country' natural reserves and protected areas. Latest estimates on naturally occurring roan

individuals in South-Africa account for about 300 individuals, with the population being considered as locally endangered (East and IUCN SSC Antelope Specialist Group 1999). Within the Limpopo province, in northeast of South-Africa, main populations occur within protected areas, such as the Kruger National Park, Nylsvley Nature Reserve and Percy Fyfe Nature Reserve. Previous assessments across these areas reported decreasing population numbers and overall low levels of genetic diversity (Grobler and Nel 1996; Harrington et al. 1999; Alpers et al. 2004; Kröger and Rogers 2005). Documented roan antelope translocations within Southern Africa dates from late 1960's to early 2000's, with animals from several regions across Botswana, Malawi, Namibia, and Zimbabwe, being transferred into South-Africa, to either found new populations, or increase the numbers for those already existing, and facing serious extinction risks (Berrie 2003). With such an intense history of animal translocation, together with increasing difficulty to maintain stable effective numbers under closed, fragmented populations, it has become crucial to assess current genetic diversity of roan antelopes across South-Africa. One of the largest natural roan populations is held at Percy Fyfe NR, founded in 1968 by individuals from the Waterberg Plateau Park, in Namibia. It is believed that many founder groups for other South-African populations were originated from Percy Fyfe NR, such as the one at Nylsvley NR. Others, such as the population of Marakele National Park, also within the Limpopo province, were more recently founded by roan antelopes from Botswana (Berrie 2003).

In the framework of my thesis, we additionally genotyped 13 contemporary samples of roan antelopes held at Marakele NP (n = 3), Percy Fyfe NR (n = 1), and Nylsvley NR (n = 9), using the 43 microsatellite loci, as previously described on Chapter 3 – Study I, under the same conditions described in section 3.2.2. Together with remaining samples from our original dataset, and after checking for relatedness, we re-analysed the genetic clustering with Structure v.2.3.4 (Pritchard, Stephens and Donnelly 2000; Falush, Stephens and Pritchard 2003; Hubisz et al. 2009), on two datasets: i) by extending the original dataset with these 13 individuals (n = 144); and ii) the same extended dataset, but including only contemporary samples (n = 73). Using the first dataset, we also analysed the genetic clustering based on Nei's genetic distances, through principal coordinate analyses (PcoA) in GeneAlEx v.6.5.3 (Peakall and Smouse 2006, 2012), and calculated summary statistics within groups, using Arlequin v.3.5 (Excoffier and Lischer 2010) and Fstat v.2.9.3 (Goudet 2001) (see Chapter 3 – section 3.2.3 for details on parameters used). Levels of differentiation between groups were calculated using the pairwise means of the fixation index (F_{ST}) and the standardized G'_{ST} statistic (Hedrick 2005).

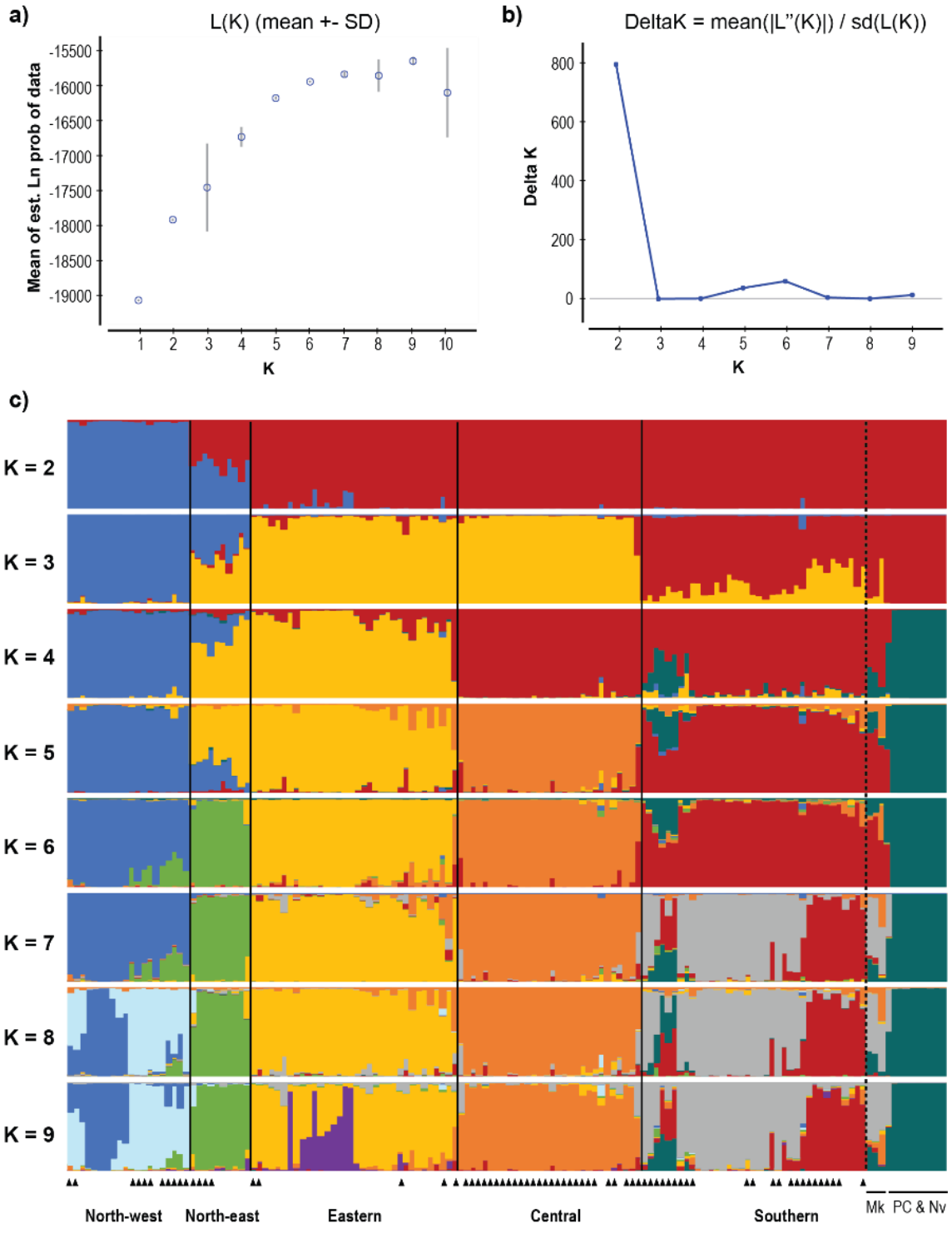


Figure 5. 4 Bayesian clustering analysis for the nuclear microsatellite data, using the extended roan antelope dataset, containing the same individuals presented in Chapter 3 – Study I, plus contemporary samples from 13 individuals across northeast South-Africa. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 9. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered as in Chapter 3 – Study I, plus individuals from Marakele NP (Mk), Percy Fyfe NR (PC), and Nylsvley NR (Nv). Triangles indicate historic samples. Full black lines divide according to groups (bold names at the bottom), as defined in Chapter 3 – Study I; dotted black line separates individuals from original dataset in Chapter 3 and the 13 added individuals.

Results from Bayesian clustering analysis on the extended dataset, from K = 2 to K = 9 (Figure 5.4), showed a coherent spatial genetic pattern, similar to the one originally retrieved in Chapter 3 – Study I, dividing a north-west, north-east, eastern, central, and southern cluster. Until K = 3, the 13 contemporary South-African samples are assigned to the southern group (*H. e. equinus* ssp.), containing remaining historic samples from South-Africa, as well as individuals from Zimbabwe, Botswana, Namibia, and Angola. However, from K = 4 onwards, individuals from Percy Fyfe NR and Nylsvley NR were assigned to its own cluster (0.6 < q < 1.0), separated from the remaining southern individuals. Assignment proportions of this new cluster were also retrieved for historical samples from South-Africa and Zimbabwe, although at lower frequency (0.1 < q < 0.4). Curiously, the genetic signature of Marakele NP roan antelopes assigned them to the southern cluster (0.6 < q < 0.9) (Figure 5.4).

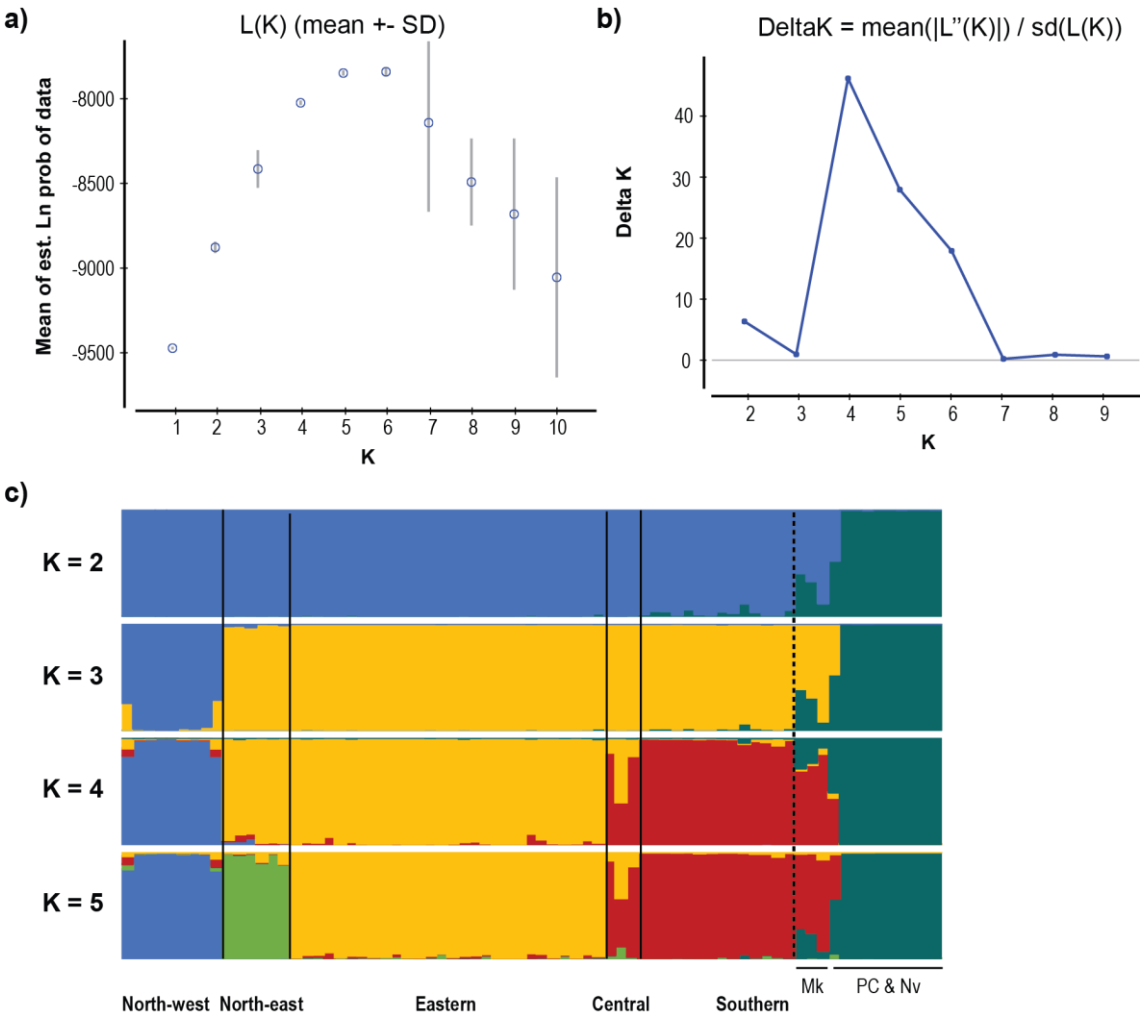


Figure 5. 5 Bayesian clustering analysis for the nuclear microsatellite data, using the contemporary samples from the extended roan antelope dataset, containing the same individuals presented in Chapter 3 – Study I, plus 13 individuals across northeast South-Africa. **a)** Posterior likelihood [L(K)] values from 10 independent Structure runs, from K = 1 to K = 10; **b)** delta-K based on the rate of change of log probability; **c)** graphs showing the representation of individual assignment

proportions (0.0 < q < 1.0), according to Structure results for K = 2 to K = 5. Individuals are represented by a vertical bar and each colour symbolizes a different cluster. Individuals ordered as in Chapter 3 – Study I, plus individuals from Marakele NP (Mk), Percy Fyfe NR (PC), and Nylsvley NR (Nv). Full black lines divide according to groups (bold names at the bottom), as defined in Chapter 3 – Study I; dotted black line separates individuals from original dataset in Chapter 3 and the 13 added individuals.

When performing the same analysis using only the contemporary samples only, individuals sampled at Percy Fyfe NR and Nylsvley NR emerge as a different cluster immediately at K = 2, divided from all the remaining ones, including the North-west individuals (Figure 5.5), whereas Marakele NP individuals were genetically assigned to the southern cluster.

The PcoA (Figure 5.6) retrieved a similar pattern to both clustering analyses, with a clear division of Nylsvley NR individuals from remaining ones; however, in this analysis, the Percy Fyfe NR individuals are genetically closer to the southern cluster.

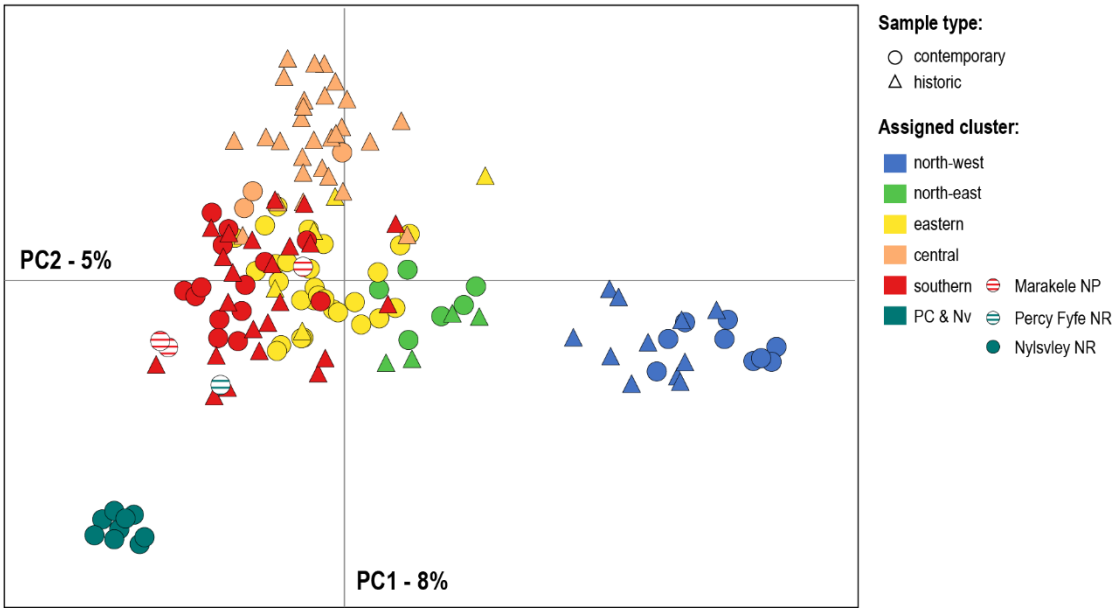


Figure 5. 6 Scatterplot of the first two principal components from PcoA results, analysing the nuclear microsatellite data, using the extended roan antelope dataset, containing the same individuals presented in Chapter 3 – Study I, plus contemporary samples from 13 individuals across northeast South-Africa. Colours correspond to the main six groups identified from the hierarchical clustering analyses (K = 6 - Figure 5.4c), according to the legend. Individuals from contemporary South-African populations (Marakele NP, Percy Fyfe NP, Nylsvley NR) are depicted, as in the figure legend. Each point represents an individual.

From Bayesian clustering analysis on the extended dataset, and looking at the clusters defined for K = 6 (Figure 5.4), we calculated within group summary statistics. Based on expected heterozygosity (H_e), average number of alleles per locus (N_a), number of private alleles (N_{pa}), and allelic richness (AR), results indicated reduced levels of diversity for the group comprising individuals from Percy Fyfe NR and Nylsvley NR,

compared to remaining groups (Table 5.1). Also, individuals from both South-African Nature Reserves had the only negative value of inbreeding-coefficient (F_{IS}), albeit non-significant ($p \geq 0.1$), possibly indicating an excess of heterozygotes. Differentiation levels measured by pairwise F_{ST} and G'_{ST} supported significant structure ($p < 0.001$) across the range of roan antelopes (results not shown), with the highest values consistently found between the group comprising of individuals from Percy Fyfe NR and Nylsvley NR, and remaining ones ($0.189 < F_{ST} < 0.270$; $0.385 < G'_{ST} < 0.548$).

Table 5. 1 Measures of nuclear genetic diversity from the analysis of 43 microsatellite loci for the extended roan antelope dataset, containing the same individuals presented in in Chapter 3 – Study I, plus contemporary samples from 13 individuals across northeast South-Africa. Results according to the six groups identified from the hierarchical clustering analyses (K = 6 - Figure 5.4c), the first five corresponding to previous ones analysed in Chapter 3 – Study I (with the southern group also containing individuals from Marakele NP); the sixth group containing individuals from Percy Fyfe NR and Nylsvley NR (PC & Nv).

	N	H_e (± s.d.)	N_a (± s.d.)	N_{pa}	AR	F_{IS}
North-west	20	0.601 ± 0.331	6.5 ± 4.9	51	5.0	0.107
North-east	10	0.580 ± 0.305	4.7 ± 2.7	12	4.5	0.063
Eastern	33	0.528 ± 0.330	6.0 ± 4.2	15	4.3	0.019
Central	31	0.503 ± 0.327	5.2 ± 3.5	9	3.8	0.059
Southern	40	0.519 ± 0.331	6.5 ± 4.8	21	4.3	0.112***
PC & Nv	10	0.316 ± 0.284	2.3 ± 1.3	2	1.9	-0.122

N, sample size; $\mathbf{H_e}$, expected heterozygosity and respective standard deviations (**s.d**); $\mathbf{N_a}$, average number of alleles per locus and respective standard deviations (**s.d**); $\mathbf{N_{pa}}$, number of private alleles; **AR**, allelic richness; $\mathbf{F_{IS}}$, inbreeding coefficient with *** $p \leq 0.001$, after Bonferroni corrections. All parameters were estimated allowing 5% of missing data per locus.

Overall, the presented results of genetic clustering, diversity, and differentiation, retrieved different genetic signatures across the three South-African natural populations. Individuals held at Percy Fyfe NR and Nylsvley NR were assigned to a different cluster from remaining southern roan antelopes (considering both current and historic range distribution), with signs of reduced genetic diversity and increased heterozygosity. This cluster has the typical genetic signature of a founder effect, artificially created in this case, with strong signs of genetic drift (Templeton 1980; Arcos-Burgos and Muenke, 2002; Pardo et al. 2005). Such results are in agreement with previous studies on these populations (Grobler and Nel 1996; Dörgeloh 2001). Also, genetic clustering seems to reflect the founding history of these populations, with the Nylsvley NR individuals being closely related to the one from Percy Fyfe NR, their source population (Berrie 2003). On the contrary, individuals held at Marakele NP exhibit a genetic signature typical of a southern roan antelope (*H. e. equinus*), expected given the recent foundation of this population, in 2001. Reduced genetic diversity from founder effects, caused by translocation of roan antelopes into natural populations, together with reported

outbreeding depression, due to intraspecific hybridization, within private land (van Wyk et al. 2019), can increase the extinction risk of roan South-African populations, already considered as locally endangered (East and IUCN SSC Antelope Specialist Group 1999). Similar conclusions have been also reported for other South-African populations of antelopes with comparable threats, such as the bontebok and blesbok, as well as, the blue and black wildebeest (van Wyk et al. 2013; Grobler et al. 2018). Even though this study holds important conclusions towards management plans across Southern African roan antelopes, these should be considered with caution, given the reduced number of sampled individuals. Expanded surveys are needed to fully address this issue, not only across South-Africa, but also remaining endangered populations across the eastern and southern distribution of the species.

5.3 Importance of natural history collections

Natural history collections (NHC), namely those housed in Museums, are reservoirs for the study of biodiversity, consisting of several types of preserved biological materials from specimens worldwide (Suarez and Tsutsui 2004; Wandeler, Hoeck and Keller 2007; Bradley et al. 2014). Their role in supporting scientific research has long been recognized, with samples contributing to several fields, such as taxonomy and systematics, ecology and conservation, as well as evolutionary biology and genomics (Smith and Blagoderov 2012; Holmes et al. 2016). With the advent of molecular approaches, the use of historical samples to study natural populations increased significantly (Bi et al. 2013; Burrell, Disotell and Bergey 2015; Lopez et al. 2020), particularly for mammal species (McLean et al. 2016). NHC enables the study of endangered or even extinct taxa, as recently reported for the *Hippotragus* genus, with the endangered Giant sable antelope (*H. niger variani*) (Vaz Pinto 2018) and the extinct bluebuck (*H. leucophaeus*) (Hempel et al. 2021). For this thesis, we obtained NHC samples to complement the extensive dataset across the pan-African distribution of the roan antelope. These collections provided samples that would be extremely difficult to obtain in the wild, due to political and/or logistical constraints. To the best of our knowledge, the data presented in this thesis are the first for several roan antelope populations across areas recently ravaged by war, such as CAR, Sudan, and Mozambique. Therefore, NHC samples were essential for this thesis, complementing the roan antelope dataset and allowing to achieve the proposed objectives.

Our dataset on the roan antelope comprised of historical samples donated from four European and one African NHC. One practical constraint when working with such samples, is related to their acquisition. After an initial contact, all European institutions

promptly sent us the requested samples. On the contrary, the African institution required for *in loco* extraction, due to lack of resources and qualified personnel. I was fortunate to belong to a research team that had the necessary funds to allow me to travel to Zimbabwe and assist on sampling collection and clearances filling at the Bulawayo Natural History Museum. However, for other projects, this fact may add logistical and/or financial constraints during sampling steps, which from my point of view, may hinder a more intense use of African NHC towards research studies.

Due to their historical nature, DNA extracted from NHC samples presents high levels of degradation, being suitable to external contamination, and therefore requires specific laboratory conditions, techniques, and qualified technical staff (Wandeler et al. 2007). The manipulation and analyses of historical samples must be performed on a dedicated laboratory for low quality DNA, incorporating special features, and be exclusively used for this type of samples. During laboratory work performed in this thesis, each technique (e.g. DNA extraction, DNA amplification, or library preparation) was staged in physically separated rooms, to avoid cross-contamination throughout the process. These ultimately affect the successful rate of analysing historical samples, but other factors, such as the type of tissue sampling and the preservation method, are certainly also relevant (Wandeler et al. 2003; Pääbo et al. 2004; Rohland and Hofreiter 2007). Because of this, sampling effort can vary considerably.

The present work was based on historical samples from 145 roan antelope individuals. Prior to laboratory work, about 10% of these (n = 15) were excluded from the dataset due to incomplete record on sampling location; such constraints are frequent when dealing with NHC samples (Wandeler et al. 2007). These historical samples covered several biological materials – tooth, bone, skull, and skin tissue. For DNA extraction, skin tissue was obtained directly from donated samples, whereas remaining types of samples required additional equipment, namely a diamond drill and a bone mill, which causes irreversible damage to the specimen (Wandeler et al. 2007). Following the Dabney et al. (2013) protocol, we performed triplicates of DNA extraction for each sample, to assure enough final eluate to perform subsequent analyses. Depending on the molecular marker, the successful rate of historical samples can also vary (Burrell et al. 2015). In this thesis, by following several protocol adjustments during microsatellite amplification (see Chapter 3 – Text S3.2 for details) (Piggott et al. 2004), the success rate for this molecular marker was 57%, whereas for whole-mitochondrial sequencing, the technique of target sequencing capture (see Chapter 3 – section 3.2.4 for details), increased the success rate to 84%. For whole-genome sequencing, it was important to select the best-quality historical samples, in terms of DNA quantity and appropriate fragment size (Burrell et al. 2015). In our roan antelope dataset, these consisted of dried-

skin tissue donated by the Powell-Cotton Museum (Kent, UK), with averaged eluted DNA of 86 ng/µl (range between 64 and 100 ng/µl), therefore accounting for more than 70% of the historical samples used for genomic re-sequencing data (Chapter 3 – Study II and Chapter 4). Using specific laboratory protocols for library preparation and amplification, (Meyer and Kircher 2010; Dabney and Meyer 2012; Kircher, Sawyer and Meyer 2012), as well as the MinElute PCR purification kit (Qiagen, Germany) during purification steps, the successful rate of historical samples for whole-genome re-sequencing was of 65%, with an average of 96% usable sequenced reads per sample.

To sum up, numerous best-practices, protocol adjustments, and enhanced approaches are currently available to overcome the practical constraints associated with NHC samples. Apart from such constraints, it is also important to weigh in the financial and time investment towards this type of sampling scheme, usually higher than for contemporary samples.

5.4 Final considerations

The roan and the sable antelope are iconic species of sub-Saharan African savannas. They are called the horse-like antelopes due to their grazing nature and body proportions, but also their impressive beauty and symbolic ominous. Their unique evolutionary history resulted from dramatic changes that the African continent experienced throughout geological times, leading to both species current distribution, and genetic diversity. Understanding their history allows us to, not only appreciate their uniqueness in terms of evolution, but also to help ensure their continued existence, as part of the amazing ecosystem that is the African savannas. In 2016 I had the privilege of experiencing such wonder for the first time and see these amazing animals in the wild. I will always cherish such magical time, and remember the colours, sounds, smells, and the heat. The amazing picture of a powerful, and yet, so fragile ecosystem. The beauty of how everything is connected: predator and prey, rain and soil, storm and fire. How everything is perfectly balanced. When we experience it first-hand, we feel the responsibility to protect and preserve it, and give a chance for future generations to experience it as well.

5.5 References

Ansell, W. F. H., 1972. Order Artiodactyla, pp. 15–83 in *Mammals of Africa: an Identification Manual*. Smithsonian Institute Press (J. Meester and H. W. Setzer, Eds.), Washington, DC.

Alpers, D. L., Jansen van Vuuren, B., Arctander, P. and Robinson, T. J., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13: 1771–1784.

Arbogast, B. S. and Kenagy, G. J., 2008. Comparative phylogeography as an Integrative approach to historical biogeography. *Journal of Biogeography*, 28: 819–25.

Arcos-Burgos, M. and Muenke, M., 2002. Genetics of population isolates. *Clinical Genetics*, 61: 233–247.

Avise, J. C., 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36: 3–15.

Barrie, A., 2003. Translocation of roan antelope in South Africa and the effect this has had on the genetic diversity of the species. PhD thesis dissertation, University of Johannesburg, South-Africa.

Bermingham, E. and Moritz, C., 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology*, 7: 367–369.

Bertola, L. D., Jongbloed, H., Van Der Gaag, K. J., de Kniijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., et al., 2016. Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the Lion (*Panthera leo*). *Scientific Reports*, 6: 30807.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R. and Moritz, C., 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, 22: 6018–6032.

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

Bibi, F. and Kiessling, W., 2015. Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences*, 112: 10623–10628.

Bobe, R. and Eck, G., 2001. Responses of African bovids to Pliocene climate change. *Paleobiology*, 27: 1–47.

Bothma, J. du P. and van Rooyen, N., 2005. *Intensive wildlife production in southern Africa*. Van Schaik Publishers (J. du P. Bothma and N. van Rooyen, Eds.), Pretoria.

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., et al., 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, 50: 362–367.

Bradley, R. D., Bradley, L. C., Garner, H. J. and Baker, R. J., 2014. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. *BioScience*, 64: 1150–1158.

Burrell, A., Disotell, T. and Bergey, C., 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, 79: 35–44.

Chardonnet, P. and Crosmary W., 2013. *Hippotragus equinus* Roan Antelope, pp. 548–556 in *Mammals of Africa VI*. Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Chorowicz, J., 2005. The East African rift system. *Journal of African Earth Sciences*, 43: 379–410.

Cousins, J. A., Sadler, J. P. and Evans, J., 2010. The challenge of regulating private wildlife ranches for conservation in South Africa. *Ecology and Society*, 15: 28–48.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, 52: 87–94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110: 15758–15763.

DeMenocal, P. B., 1995. Plio-Pleistocene African climate. *Science*, 270: 53–59.

DeMenocal, P. B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, 220: 3–24.

Dörgeloh, W. G., 2001. A draft habitat suitability model for roan antelope in the Nylsvley Nature Reserve, South Africa. *African Journal of Ecology*, 39: 313–316.

Dupont, L., 2011. Orbital scale vegetation change in Africa. *Quaternary Science Reviews*, 30: 3589–3602.

Du Toit, J. T. and Cumming, D. H. M., 1999. Functional significance of ungulate diversity in African savannas and the ecological implications of the spread of pastoralism. *Biodiversity and Conservation*, 8: 1643–1661.

East, R. and IUCN SSC Antelope Specialist Group, 1999. *African antelope database 1998* (IUCN/SSC Antelope Specialist Group, Ed.), Gland: IUCN.

Estes, R., 2013. *Hippotragus niger* Sable antelope, pp. 556–565 in *Mammals of Africa VI.* Bloomsbury Publishing (K. J. D. Happold, M. Hoffmann, T. Butynski, M. Happold, K. Kalina, et al., Eds.), London.

Estevo, C. A., Nagy-Reis, M. B. and Nichols, J. D., 2017. When habitat matters: habitat preferences can modulate co-occurrence patterns of similar sympatric species. *Plos One*, 12: e0179489.

Excoffier, L. and Lischer., H. E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564–567.

Falush, D., Stephens, M. and Pritchard, J. K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics Society of America*, 164: 1567–1587.

Fernández, M. H. and Vrba, E. S., 2005. Rapoport effect and biomic specialization in African mammals: revisiting the climatic variability hypothesis. *Journal of Biogeography*, 32: 903–918.

Fjeldså, J., Burgess, N. D., Blyth, S. and De Klerk, H. M., 2004. Where are the major gaps in the reserve network for Africa's mammals? *Oryx*, 38: 17–25.

Goudet, J., 2001. Fstat, version 2.9.3.2. A program to estimate and test gene diversities and fixation indices. Institute of Ecology University of Lausanne, Switzerland.

Goudie, A. S., 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67: 437–456.

Grobler, J. and Nel, G., 1996. Lack of allozyme heterogeneity at eight loci studied in a roan antelope *Hippotragus equinus* population from the Percy Fyfe Nature Reserve. *South African Journal of Wildlife*, 26: 34–35.

Grobler, P., van Wyk, A. M., Dalton, D. L., Jansen van Vuuren, B. and Kotzé, A., 2018. Assessing introgressive hybridization between blue wildebeest (*Connochaetes taurinus*) and black wildebeest (*Connochaetes gnou*) from South Africa. *Conservation Genetics*, 19: 981–993.

Harrington, R., Owen-Smith, N., Viljoen, P. C., Biggs, H. C., Mason, D. R. and Funston, P., 1999. Establishing the causes of the roan antelope decline in the Kruger National Park, South Africa. *Biological Conservation*, 90: 69–78.

Hedrick, P. W., 2005. A standardized genetic differentiation measure. *Evolution*, 59: 1633–1638.

Heller, R., Lorenzen, E. D., Okello, J. B. A., Masembe, C. and Siegismund, H. R., 2008. Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Molecular Ecology,* 17*:* 4845-4858.

Heller, R., Brüniche-Olsen, A. and Siegismund, H. R., 2012. Cape buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. *Molecular Ecology*, 21: 3947-3959.

Hempel, E., Bibi, F., Faith, J., Brink, J., Kalthoff, D., Kamminga, P., Paijmand, J. L. A., Westbury, M. V., Hofreiter, M. and Zachos, F. E., 2021. Identifying the true number

of specimens of the extinct blue antelope (*Hippotragus leucophaeus*). *Scientific Reports*, 11: 1–14.

Hewitt, G., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359: 183–195.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F. and Yoder, A. D., 2010. Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*, 54: 291–301.

Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., Labarbera, K., Wommack, E. A., et al., 2016. Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25: 864–881.

Hubisz, M. J., Falush, D., Stephens, M. and Pritchard, J. K., 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9: 1322–1332.

IUCN SSC Antelope Specialist Group, 2008. *Hippotragus niger* ssp. *variani*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10169A50188611.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017a. *Hippotragus equinus*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10167A50188287.en on April 8th, 2020.

IUCN SSC Antelope Specialist Group, 2017b. *Hippotragus niger*. Retrieved from https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T10170A50188654.en on April 8th, 2020,

Jansen van Vuuren, B., Robinson, T. J., Vaz Pinto, P., Estes, R. and Matthee, C. A., 2010. Western Zambian sable: are they a geographic extension of the giant sable antelope? *South African Journal of Wildlife Research*, 40: 35–42.

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., Rubenstein, D. I., Wang, W. and Pringle, R. M., 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences*, 112: 8019–8024.

Kircher, M., Sawyer, S. and Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40: e3–e3.

Koepfli, K.-P., Tamazian, G., Wildt, D., Dobrynin, P., Kim, C., Frandsen, P. B., Godinho, R., Yurchenko, A. A., et al., 2019. Whole genome sequencing and re-sequencing of the Sable antelope (*Hippotragus niger*): a resource for monitoring diversity in *ex-*

situ and *in-situ* populations. *G3:Genes, Genomes, Genetics*, 9: 1785–1793.

Kröger, R. and Rogers, K. H., 2005. Roan (*Hippotragus equinus*) population decline in Kruger National Park, South Africa: influence of a wetland boundary. *European Journal of Wildlife Research*, 51: 25–30.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J. and Rahbek, C., 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39: 1189–1205.

Lopez, L., Turner, K. G., Bellis, E. S. and Lasky, J. R., 2020. Genomics of natural history collections for understanding evolution in the wild. *Molecular Ecology Resources*, 20: 1153–1160.

Lorenzen, E. D., Heller, R. and Siegismund, H. R., 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, 21: 3656–3670.

Mayaux, P., Bartholomé, E., Fritz, S. and Belward, A., 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31: 861–877.

McLean, B. S., Bell, K. C., Dunnum, J. L., Abrahamson, B., Colella, J. P., Deardorff, E. R., Weber, J. A., Jones, A. K., Salazar-Miralles, F. and Cook, J. A., 2016. Natural history collections-based research: progress, promise, and best practices. *Journal of Mammalogy*, 97: 287–297.

Meyer, M. and Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010: pdb.prot5448.

Miller, J. M., Hallager, S., Monfort, S. L., Newby, J., Bishop, K., Tidmus, S. A., Black, P., Houston, B., et al., 2011. Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). *Conservation Genetics*, 12: 423–431.

Moritz, C., 1994. Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, 9: 373–375.

Moritz, C., 1999. Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas*, 130: 217-228.

Ogutu, J. O., Piepho, H. P. and Dublin, H. T., 2015. Reproductive seasonality in African ungulates in relation to rainfall. *Wildlife Research*, 41: 323–342.

Orteu, A. and Jiggins, C. D., 2020. The genomics of coloration provides insights into adaptive evolution. *Nature Reviews Genetics*, 21: 461–475.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M., 2004. Genetic analyses from ancient DNA. *Annual Review of Genetics*, 38: 645–679.

Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. and Aulchenko, Y. S., 2005. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics*, 69: 288–295.

Peakall, R. and Smouse, P. E., 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6: 288–295.

Peakall, R. and Smouse, P. E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28: 2537–2539.

Piggott, M. P., Bellemain, E., Taberlet, P. and Taylor, A. C., 2004. A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics*, 5: 417–420.

Piltz, J., Sorensen, T. and Ferrie, G. M., 2016. Population analysis and breeding and transfer plan: sable antelope (*Hippotragus niger*). AZA Species Survival Plan Yellow Program. Report by *AZA Population Management Center*.

Pitman, R. T., Fattebert, J., Williams, S. T., Williams, K. S., Hill, R. A., Hunter, L. T., Slotow, R. and Balme, G. A., 2017. The conservation costs of game ranching. *Conservation Letters*, 10: 403–413.

Pitra, C., Hansen, A. J., Lieckfeldt, D. and Arctander, P., 2002. An exceptional case of historical outbreeding in African sable antelope populations. *Molecular Ecology*, 11: 1197–1208.

Pitra, C., Vaz Pinto, P., O'Keeffe, B. W. J., Willows-Munro, S., Jansen van Vuuren, B. and Robinson, T. J., 2006. DNA-led rediscovery of the giant sable antelope in Angola. *European Journal of Wildlife Research*, 52: 145–152.

Pritchard, J. K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.

Ratnam, J., Bond, W. J., Fensham, R. J., Hoffmann, W. A., Archibald, S., Lehmann, C. E., Anderson, M. T., Higgins, S. I. and Sankaran, M., 2011. When is a 'forest' a savanna, and why does it matter? *Global Ecology and Biogeography*, 20: 653–660.

Rocha, J., 2014. The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences. MsC thesis dissertation, University of Porto, Portugal.

Rohland, N. and Hofreiter, M., 2007. Ancient DNA extraction from bones and teeth. *Nature Protocols*, 2: 1756–1762.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., et al., 2005. Determinants of woody cover in African savannas. *Nature*, 438: 846–849.

Smith, V. and Blagoderov, V., 2012. Bringing collections out of the dark. *ZooKeys*, 209: 1–6.

Smitz, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., Caron, A., Prins, et al., 2013. Pan-African genetic structure in the African Buffalo (*Syncerus caffer*): investigating intraspecific divergence. *Plos One*, 8: e56235.

Suarez, A. and Tsutsui, N., 2004. The value of museum collections for research and society. *BioScience*, 54: 66–74.

Templeton, A. R., 1980. The theory of speciation via the founder principle. *Genetics*, 94: 1011–1038.

Thomassen, H. A., Freedman, A. H., Brown, D. M., Buermann, W. and Jacobs, D. K., 2013. Regional differences in seasonal timing of rainfall discriminate between genetically distinct East African giraffe taxa. *Plos One*, 8: e77191.

Trauth, M. H., Larrasoaña, J. C. and Mudelsee, M., 2009. Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, 28: 399–411.

van Beest, F. M., McLoughlin, P. D., Vander Wal, E. and Brook, R. K., 2014. Density-dependent habitat selection and partitioning between two sympatric ungulates. *Oecologia*, 175: 1155–1165.

van Wyk, A. M., Kotzé, A., Randi, E. and Dalton, D. L., 2013. A hybrid dilemma: a molecular investigation of South African bontebok (*Damaliscus pygargus pygargus*) and blesbok (*Damaliscus pygargus phillipsi*). *Conservation Genetics*, 14: 589–599.

van Wyk, A. M., Dalton, D. L., Kotzé, A., Grobler, P. J., Mokgokong, P. S., Kropff, A. S. and Jansen van Vuuren, B., 2019. Assessing introgressive hybridization in roan antelope (*Hippotragus equinus*): lessons from South Africa. *Plos One*, 14: e0213961.

Vaz Pinto, P., 2009. Giant sable rescue and translocation. *Gnusletter*, 28: 8–10.

Vaz Pinto, P., 2018. Evolutionary history of the critically endangered giant sable antelope (*Hippotragus niger variani*). Insights into its phylogeography, population genetics, demography and conservation. PhD thesis dissertation, University of Porto, Portugal.

Vrba, E., 1996. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate, pp. 385–424 in *Paleoclimate and Evolution*. Yale University Press (E. Vrba, G. Denton, T. Partidge and L. Burckle, Eds.), Connecticut.

Wandeler, P., Smith, S., Morin, P. A., Pettifor, R. A. and Funk, S. M., 2003. Patterns of nuclear DNA degeneration over time - a case study in historic teeth samples. *Molecular Ecology*, 12: 1087–1093.

Wandeler, P, Hoeck, P. and Keller, L., 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, 22: 634–642.

Wildt, D., Miller, P., Koepfli, K.-P., Pukazhenthi, B., Palfrey, K., Livingston, G., et al., 2019. Breeding centers, private ranches, and genomics for creating sustainable wildlife populations. *BioScience*, 69: 928–943.

Winter, S., Fennessy, J. and Janke, A., 2018. Limited introgression supports division of giraffe into four species. *Ecology and Evolution*, 8: 10156–10165.