

Assessing the genetic diversity and structure of spotted hyena (*Crocuta Crocuta*) populations across Africa

By

Kagisho Dibakoane

Thesis presented in fulfilment of the requirements for the Degree of Master of
Science in Biological Sciences: Zoology

Supervisor: Prof. Yoshan Moodley

Co-Supervisor: Dr. Michael Vincent Westbury

Department of Biological Sciences

Faculty of Science, Engineering and Agriculture

University of Venda

Abstract

The spotted hyena (*Crocuta crocuta*) is one of the most important species in the savannah ecosystem acting as both a predator and a scavenger yet not a lot is known about this majestic species. This study aims to address the genetic diversity, population structure and gene flow of these species across different regions of the sub-Saharan distribution. 55 whole genome and restriction site-associated sequences were analysed using Heterozygosity, PCA analysis, Admixture proportions, Phylogenetic analysis and D-statistics. The results showed that the mean heterozygosity of spotted hyena was 0.00094 (RAD-seq) and 0.0011 (WGS) ranging from 0.0006 to 0.0013. Admixture proportions and phylogenetic analyses showed three main populations clusters (western, eastern and southern) but there was also detailed structuring which showed five populations (west, northeast, eastern TanzaniaKenya, westernTanzania and southern) and D-statistics showed that there is gene flow between these populations. The spotted hyena shows highly structured clusters between the western, eastern and southern, this pattern of structure was comparable to other mammals and large carnivores in the same distribution. Although there was evidence of gene flow between these regions the western appeared more differentiation from either the eastern or southern both of which shared considerable genetic material with each other was also the western showed the lowest genetic diversity between these regions, these structures are cause by historical events and anthropogenic factors. We suggest that conservation effects must focus on fostering the exchange of genetic material from the eastern and southern regions to the western regions as a priority to avoid further differentiation.

Declaration

I, Kagisho Dibakoane, University of Venda student number 16017847 declare that this research dissertation is my work and has not been submitted for a degree at any other university or institution.



.....

Signature (Student)

Acknowledgements

I would like to thank my supervisors, Prof. Yoshan Moodley and Dr. Michael Westbury, for their guidance and support throughout my Master's project. I also want to acknowledge the financial support from NRF (National Research Fund) (MND210527604271) that made this research possible.

Table of Contents

Abstract.....	i
Declaration.....	ii
Acknowledgements.....	iii
1. Introduction	1
1.1. Literature review on the spotted hyena.....	1
1.1.1. Evolutionary history	1
1.1.2. Population Ecology of <i>Crocuta crocuta</i>	2
1.2. Genetic diversity of spotted hyenas in Africa	4
1.3. Population structure of spotted hyenas in Africa	6
1.4. Justification of the study.....	8
1.5. Statement of the problem.....	9
2. Methods	11
2.1. Sample details.....	11
2.2. Reference and Outgroup	12
2.3. Data processing	13
2.3.1. Mapping	13
2.3.2. Quality of the mapped data	14
2.3.3. Autosomal and sex-linked scaffold identification	14
2.3.5. Relatedness	14
2.4. Diversity	16
2.4.1. Heterozygosity.....	16
2.5. Population structure analyses	16
2.5.1. Principal Component Analysis.....	16
2.5.2. Admixture Analysis	17
2.5.3. Phylogenetic analysis	17
2.6. Gene flow.....	18
3. Results	19
3.1. Coverage.....	19
3.2. SNPs data	22

3.3. Relatedness	22
3.4. Heterozygosity.....	24
3.5. Population Structure	25
3.5.1. Principal Component Analysis	25
3.5.2. Admixture proportions.....	26
3.5.3. Phylogenetic Analysis.....	29
3.6. Gene flow.....	30
4. Discussion.....	32
4.1. Genetic diversity.....	32
4.2. Population Structure	34
4.3. Gene flow.....	36
4.4. Conservation Implications	37
References.....	38
Supplementary Information	45
Appendix A:	45
Appendix B:	47
Appendix C:	48
Appendix D:	49

1. Introduction

1.1. Literature review on the spotted hyena

1.1.1. Evolutionary history

Hyaenids first appeared in the Oligocene period around 29 mya (million years ago) and were most diverse in the Miocene, 23 to 5 mya; more than 80 fossil species have been recorded across Europe, Asia, and Africa (Turner *et al.*, 2008). The hyaenid species have been categorized into six ectomorph groups. The first group includes civet-like insectivore/omnivore that consists of the genus *Proticittherium*, which is known from deposits in Europe (Werdelin & Solounias, 1991). The second group, mongoose-like insectivore/omnivores, consists of a single genus, *Plioviverrops*. The next category comprises jackal and wolf-like meat and bone eaters, which includes three genera: *Thalassictis*, *Hyaenictitherium*, and *Ictitherium*. Following that are the cursorial meat and bone eaters, represented by three genera: *Lycyaena*, *Hyaenictis*, and *Chasmaporthetes*. The transitional bone crackers group has four genera: *Belbus*, *Palinhyaena*, *Leecyaena*, and *Ikelohyaena*. Finally, there are the fully developed bone crackers, which include *Adcrocuta*, *Pliocrocuta*, *Pachycrocuta*, and *Crocuta* (Turner *et al.*, 2008).

Hyaenids started to decline in distribution and diversity from the late Miocene. Today, this carnivore family only has four extant members, which include the spotted hyena (*Crocuta crocuta*), striped hyena (*Hyaena hyaena*), aardwolf (*Proteles cristatus*) and brown hyena (*Parahyaena brunnea*) (Werdelin & Solounias, 1991). The aardwolf is the most differentiated from the other hyaenids as it specialises in insectivory, due to the dental characterisization leading towards the omnivore/insectivore ectomorph. Turner (2008) suggested that the aardwolf may have evolved from the *Plioviverrops* in contrast to the other hyaenids that are labelled bone-cracking hyenas due to their dentition and specialising in hypercarnivory (Werdelin, 1989). The bone-cracking ecomorph taxa in this group showed an increase in size; these four species have very few key characteristics that distinguish them from each other. The genus *Crocuta* has rounded ears compared to *Hyaena*'s pointed ears. The *Crocuta* lacks a dorsal mane (Hayssen & Noonan, 2021) while present in *Hyaena* and the *Crocuta* has a spotted pelage while the *Hyaena* has a striped and uniform pelage (Ellerman *et al.*, 1953); the spotted hyenas are known to be sympatric to the brown hyena in southern Africa and striped hyena in eastern Africa (Kruuk, 1972).

The genus *Crocota* was distributed across Africa and Eurasia, with fossil evidence indicating that *C. crocuta spelaea* was found in cave sites in Europe and *Crocota ultima* was seen in Asia around one million years ago known as cave hyenas (Kalmykov & Kobylkin, 2017; Sheng *et al.*, 2013). Cave hyenas were described as the subspecies of spotted hyena and are estimated to have split around 1.85 mya. Cave hyenas still share haplogroups with spotted hyenas, although morphological characteristics distinguish spotted hyenas from cave hyenas. For instance, the cave hyena had shorter distal limbs than the spotted hyena, the metastyloid's elongation in the spotted serves as evidence of increased cutting ability compared to the cave hyena, and the cave hyenas had a less developed brain (Ellerman *et al.*, 1953). However, there is evidence of bidirectional gene flow between the extinct cave hyena and the extant spotted hyena after the divergence (Rohland *et al.*, 2005; Westbury *et al.*, 2022).

During the late Pleistocene, several large-scale events occurred, such as the megafauna extinction, *Homo erectus* migration out of Africa and the climatic transitions from the Pleistocene into the Holocene period. These events caused the extinction of large mammals, including the Neanderthal (Varela *et al.*, 2010). The *Crocota* was one of the mammals that became extinct in Eurasia in the late Pleistocene and subsequently became restricted in Africa. The evolution of the genus *Crocota* is mostly unresolved in Africa, initially most taxa were mainly confined into *Crocota crocuta*. There are findings that show that there were other taxa in Africa, such as the *Crocota ultra* and *Crocota dietrichi* (Werdelin & Lewis, 2008)

1.1.2. Population Ecology of *Crocota crocuta*

The spotted hyenas are one of the most known large mammals in Africa, occupying most of the sub-Saharan distribution (Angola, Botswana, Chad, Central African, Ethiopia, Ghana, Kenya, Tanzania, Namibia, South Africa, and South Sudan) and can be found in diverse habitats such as grasslands, savannas, woodlands, mountain tops sub-deserts and forest edges (Kruuk, 1972). Figure 1 shows the extant and extinct range of the spotted hyena population with the species being completely extinct in Europe and Asia. Spotted hyenas are estimated to be between 27,000 - 47,000 individuals, with the largest populations found in the Serengeti (Tanzania and Kenya) and the southern regions of Africa (Vissia *et al.*, 2021). However, small, stable populations can be found in the protected areas in the southern and eastern regions of Africa (Trinkel, 2009).

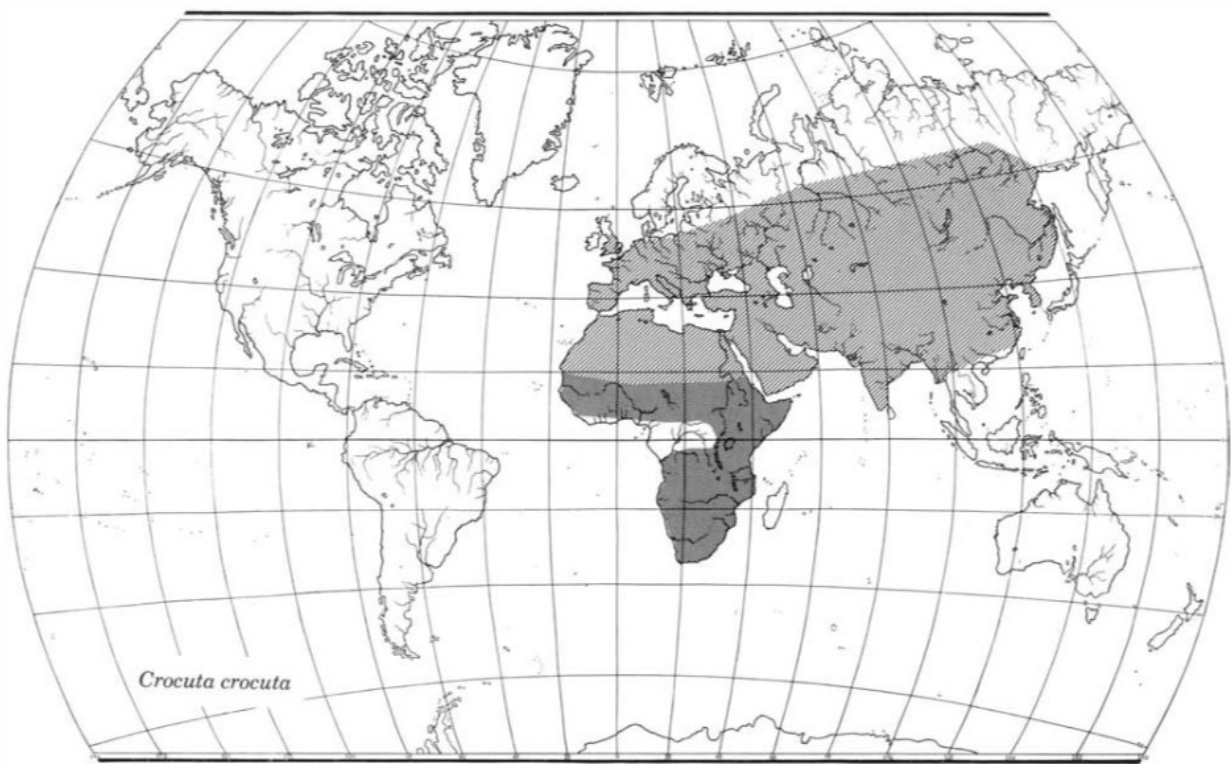


Figure 1: The range of extinct and extant populations of spotted hyena (Dark grey = extant and Light grey = extinct), adopted from Werdelin & Solounias, 1991. This figure illustrates the drastic reduction of spotted hyena populations over time, the extant populations are currently distributed across sub-Saharan Africa.

Spotted hyenas have robust teeth that enable them to crush bones, and unlike any other carnivore, they can digest all organic components (Werdelin, 1989; Palmqvist *et al.*, 2011). They hunt prey such as cape buffalo (*Syncerus caffer*), impala (*Aepyceros melampus*), wildebeest (*Connochaetes taurinus*), Thomson's gazelles (*Gazella thomsonii*), and zebra (*Equus quagga*). The spotted hyenas are also known for their distinct dietary niches as scavengers and predators; they play an essential role as carrion consumers in the ecosystem food chain (Kruuk, 1972).

Spotted hyena density is correlated to prey density and their reproduction rate increases when there is a high prey availability, the highest spotted hyena density in Africa was recorded in Botswana (Okavango Delta) and South Africa (uMkhuze Game Reserve), notably due to the high density of

impalas (Cozzi *et al.*, 2013; De Blocq, 2014; Vissia *et al.*, 2021). Although spotted hyenas are also known to survive in regions with low prey density, such as Ethiopia, where they coexist with people and primarily play a role as carrion consumers by feeding on anthropogenic foods (Yirga *et al.*, 2013), this also highlights the species' plasticity and ability to adapt to harsh environments.

Spotted hyenas live in groups called clans (6-90 individuals). They have a more socially complex structure than any other carnivore (Holekamp & Dloniak, 2010), other studies show that the social system of spotted hyenas is more comparable to that of primates based on their communication, large group sizes, hierarchical organisation, and high social intelligence (Holekamp *et al.*, 2012). Female spotted hyenas tend to spend their lives living with the natal clan, whereas male hyenas usually leave their natal clan when they reach maturity (Boydston *et al.*, 2002). The female spotted hyenas are more dominant than the males within clan; females have uniquely masculinised external genitalia and are larger than the males (Kruuk, 1972). These characteristics give the females the advantage of selecting breeding mates (Engh *et al.*, 2002).

Spotted hyena clans generally have natal females, their offspring, and immigrant males who are mostly preferred by the females for breeding (Engh *et al.*, 2002; Honer *et al.*, 2007). Furthermore, patterns of female mate preference indicate low reproductive skew among immigrant males within the clans (Holekamp *et al.*, 2007) suggesting that within a clan it is usually the males that are forced to migrate into new clans. Therefore, male spotted hyenas have shown increased rates of dispersal over females, can move over large distances (Hofer & East, 1993), and are adapted to move across potential barriers, such as regions of human settlement, due to their behavioural plasticity. The social behaviour of the spotted hyena may influence population dynamics and persistence.

1.2. Genetic diversity of spotted hyenas in Africa

Genetic diversity is the variation of alleles and genotypes within a population, and it provides the foundation for adaptation and evolutionary potential (Frankham *et al.*, 2002). Variation within populations is essential for a population's survival and reproductive success. When populations have a higher proportion of individuals with homozygous alleles or lack of heterozygous alleles, it indicates that there are fewer fit individuals in that population and will have a higher probability of undergoing inbreeding depression (Ellegren & Galtier, 2016).

Genetic diversity can be both a cause and a consequence of ecological processes (Vellend & Geber, 2005); the variable nature of gene and genome sequences has been identified as a critical aspect of biological evolution. For example, genetic diversity allows prey populations to evolve, which affects predator population dynamics. Thus, genetic diversity further drives ecological and evolutionary changes within the predator population, leading to predictable predator-prey eco-evolutionary dynamics (Garner *et al.*, 2005).

Genetic diversity is a primary source of biodiversity, and the World Conservation Union has recognised the maintenance of genetic diversity as one of the three global conservation priorities because genetic diversity is required for quantifying variability, evaluating losses, and estimating the success of conservation efforts (Frankham *et al.*, 2002). Populations that have low genetic diversity are usually, populations that are small or have limited connectivity to other population hence the gene pool becomes reduced. Such reduction increases the chance of inbreeding, which could lead to bottlenecks and extinction (Ceballos *et al.*, 2017).

In this study, the genetic diversity of spotted hyena populations in Africa will be used to quantify the variability within populations; it is widely known that genetic diversity can be predicted through life history rather than population history since genetic polymorphism carried by any species is shaped by the demographic history of that species (Vellend & Geber, 2005). Therefore, it can be expected that the demographic fluctuations due to ecological drivers such as climate change, pest outbreaks and human activities will affect genetic diversity (Ellegren & Galtier, 2016). Studies that compared the modern mammal species to their Pleistocene counterparts or family members have shown a loss of genetic diversity over time, with few examples being species such as the muskox (*Ovibos moschatus*) and bison (*Bison bison*), which went extinct, like the spotted hyena, in Eurasia. Although muskoxen and bison are becoming restricted in North America, and spotted hyenas in Africa are facing similar challenges (Shapiro *et al.*, 2004; MacPhee *et al.*, 2005; Hofreiter & Barnes, 2010), this situation is particularly concerning because the genetic diversity of spotted hyena populations across Africa has not been extensively studied.. Demographic inferences show a decline in the adequate population size of spotted hyenas over the last million years (Macphee *et al.*, 2005), and these populations have a similar demographic history across Africa (Westbury *et al.*, 2022).

The dispersal mechanism of spotted hyenas can be associated with inbreeding avoidance since males disperse to neighbouring clans, and male dispersal has been described as a critical factor in the maintenance of genetic diversity in the Amboseli population found in Kenya (Watts *et al.*, 2011). It is known that female spotted hyenas prefer to mate with immigrant males rather than males from their clans (Van Horn *et al.*, 2008). A long-term study suggests that females do not usually produce

paternally related offspring, selecting new partners to breed with every season (Szykman *et al.*, 2007; Holekamp *et al.*, 2012). Species that have structured social systems often exhibit non-random mating patterns and biased dispersal based on sex thus allowing them to have larger effective population sizes and maintain elevated levels of genetic diversity than species which lack social structuring (Snug *et al.*, 1996). The Amboseli and Masai Mara populations showed similar levels of genetic diversity (Watts *et al.*, 2011). Hence, the levels of genetic diversity in spotted hyenas will be similar across the sampled populations.

1.3. Population structure of spotted hyenas in Africa

The genetic population structure of a species is formed by the pattern of distribution of genetic variation within and between populations. Species are not genetically homogeneous over a geographical range but rather structured into groups of individuals that are restricted from one another, whether by social behaviour, environmental or anthropogenic factors (Woodruff, 2004). These mechanisms can limit the exchange of reproducing individuals between groups, resulting in genetic differences between the groups, thus causing a pattern of genetic variation between groups over a geographical range (Laikre, 2005). Assessing the genetic population structure of a species requires genetic markers such as allozymes, mitochondrial DNA, microsatellites, or single nucleotide polymorphisms. Neutral markers are preferred for analysing population structure because they show the evolutionary processes of mutation, random genetic drift, and gene flow and how these processes affect the whole genome.

In this study, the genetic population structure of spotted hyena populations across the sub-Saharan range will be assessed to understand the general patterns of genetic variation of this species across Africa. Studies on the genetic structuring of spotted hyena populations have been minimal. There is a study using mitochondrial DNA sequences from 18 individuals that found that the spotted hyenas populations show a strong phylogeographical pattern, whereby there is a split between the south and northern populations of spotted hyenas with limited geographic overlap (Rohland *et al.*, 2005). An mtDNA study of the spotted hyena populations indicated two separate haplogroups with one forming a northern clade (Ghana) and the other forming a south-eastern clade (Namibia, Botswana, Zambia, Tanzania and Kenya). This geographical split between the northern (west and central Africa) and southern (south and east Africa) populations was also visible in other mammals that shared the same geographical range, such as the African lion (*Panthera leo*), roan (*Hippotragus equinus*), warthog (*Phacochoerus africanus*), and hartebeest (*Alcelaphus buselaphus*) (Smitz *et al.*, 2018;

Ouma *et al.*, 2010; Muwanika *et al.*, 2003; Matthee & Robinson, 1999; Lorenzen *et al.*, 2012). Conversely, a study of the African leopard shows no phylogeographical split between the north and south populations (Pečnerová *et al.*, 2021).

Even fewer studies assess the genetic variation within groups, with only one known study by Watts *et al.*, (2011) focusing on the populations found in Kenya using microsatellite markers, which indicated significant structuring between the Amboseli and Mara populations. A master's thesis by Mertens De Vry (2021) used RAD-seq (Restriction site-associated DNA sequencing) data from 70 individuals to create an SNP (Single Nucleotide Polymorphisms) panel for spotted hyena populations found in Zambia; this is the only known study that utilises RAD-seq data acquired from the spotted hyenas. In the study, a PCA (Principal Component Analysis) plot was produced that showed a clear separation between the populations found in Zambia (Liuwa and Luangwa). This separation was mainly caused by the Zambezi river forming a geographical barrier but there are also anthropogenic factors such as the Capital city of Zambia reducing the connectivity between Kafue and Liuwa. These two studies raised a concern about the genetic structure of spotted hyenas within regions (Mertens De Vry, 2021), because limited gene flow between regions could cause genetic differentiation between population and later reproductive isolation may occur if they restrictions are not addressed.

A paleogenomics study by Westbury *et al.* (2020) combined paleogenomes with modern genomes from 32 individuals, 10 of which were whole genomes of modern spotted hyenas sampled from across Africa. The study revealed strong divergence between African and Eurasian populations. It should be noted that the Northern (Ghana and Somalia) modern spotted hyena clusters with the European ancient genomes suggesting that the northern lineage could be the descendent of the European ancient genomes and the same separation is seen on the Rohland study (Rohland *et al.*, 2005). Additionally, this remains the only study to look at the whole genome samples of spotted hyenas in Africa, although the samples were not sufficient for conclusive population inferences about the spotted hyena in Africa, but from these previous studies, it can be assumed that there is a genetic differentiation between the northern and southern populations. This study will address the population structure of African spotted hyena populations between geographical regions and presumed populations within regions. Based on previous studies, I predict that there will be significant structuring in populations sampled between regions, especially between the northern and southern populations and within regions.

1.4. Justification of the study

Several threats currently affect the biodiversity of wildlife populations, such as the loss of viable habitat and alterations, and increased mortality rates from exploitations, invasive species, infectious diseases, the challenges faced from climate change, and human-caused pollution. Therefore, it is essential to understand the threats that affect wildlife and how we can use genomic tools to address the significant challenges faced by many animal species in wildlife conservation. These include assessing populations and species that are overlooked when considering policies for conservation and assessment of population sizes and connectivity to better manage and conserve biodiversity (Hohenlohe *et al.*, 2021).

Genomics studies have given us tools that can be used to estimate the fundamental characteristics of wildlife populations that are important for conservation strategies, such as inbreeding, effective population size, population structure, and demographic history. Population genomic studies are typically used to identify genetic loci and variants that are commonly known to cause adaptations to evolving environments and climate, thus ensuring that conservation efforts can be amended or invented to analyse the capacity of populations to evolve and adapt in response to those changes (Hohenlohe *et al.*, 2021).

Previous genetic studies regarding populations of spotted hyenas used microsatellite markers. Recently, SNP markers have been presented as better genetic markers since they have lower mutation rates than microsatellites (Schlötterer, 2004). The bi-allelic nature of SNPs (single-nucleotide polymorphisms) is more accurate due to the simplified genotyping process. Lastly, microsatellites are less suitable for producing high-quality genotypes from low-quality or fragmented DNA because microsatellites (80-300 bp) require longer entire sequences for amplification compared to SNPs (50-70 bp) (Morin & McCarthy, 2007). Although SNPs can provide a lower resolution and statistical power than microsatellites, this can be overcome by increasing the number of SNP markers used (Selkoe & Toonen, 2006).

Latest genomic studies utilise whole-genome sequences (Westbury *et al.*, 2020; Yang *et al.*, 2020; Shao *et al.*, 2022), but for population studies, sequencing hundreds of individuals is more costly compared to RAD-seq. Therefore, RAD-seq is more advantageous in population studies due to the low cost of sequencing multiple individuals. Still, it also comes at a disadvantage because the reduced representation libraries depend on the restriction enzymes and PCR steps (Shafer *et al.*,

2017). More errors are bound to occur mainly if a mutation occurs at an enzyme cut site on an individual that will cause allelic dropout and can lead to biased inferences (Cariou *et al.*, 2016).

1.5. Statement of the problem

There needs to be more knowledge of the genetic variation of African spotted hyena populations compared to their relatives and other mammals, which share the same habitats. For instance, the brown hyena, which is sympatric to the spotted hyena in the southern parts of Africa, has shown a decline in population size, resulting in a loss of genetic diversity of the species (Westbury *et al.*, 2022). For species that once occupied Eurasian areas and then became restricted in Africa, loss of genetic diversity is expected in such declining populations; such reduction in genetic diversity has been recorded in the North American bison, which occupied Eurasia (Macphee *et al.*, 2005).

The phylogeography of spotted hyenas still needs to be better understood, although evidence suggests a north and south subdivision (Rohland *et al.*, 2005). The eastern (Kenya & Tanzania) and southern (Namibia, Botswana & Zambia) genetic patterns of spotted hyena populations are still unknown, and the genetic differentiation of populations in Kenya leads us to wonder about the genetic differentiation of populations in other regions. So far, there still needs to be a population study for other populations; hence, this study aims to address the genetic differentiation between populations and within populations in Botswana, Namibia, Zambia, Tanzania, Kenya, Somalia, and Ghana. Although a north-south subdivision is known, it is not tested at the population level with NGS data and a more substantial number of samples that include areas that were not sampled before. This study will reveal the population structure of spotted hyena populations on the continental and regional scale.

Research Questions

- 1 How does the genetic diversity of spotted hyena populations compare in different regions and to other carnivores in the same guild such as the African lion, African leopard, Cheetah and the African wild dog?
- 2 What is the phylogeographical pattern of spotted hyena populations from the south, east and western regions?
- 3 What is the level of connectivity between geographical regions and presumed populations within regions?

Objectives

Broad Objective

This study uses RAD-seq data to assess the genetic diversity, structure, and gene flow of spotted hyena populations.

Specific Objectives

1. To assess the heterozygosity of spotted hyenas between populations in Africa.
2. To determine the population structure of spotted hyena populations in different regions in Tanzania (Selous, Maswa, Ugalla, Monduli, Kigozi, Arusha & Rungwa), Kenya_(Marsbit, Amboseli & Masai Mara), Ghana (Dusie, Jang, Gushiegu and Kabampe), Botswana and Namibia.
3. To analyse gene flow between populations in different regions Tanzania (Selous, Maswa, Ugalla, Monduli, Kigozi, Arusha & Rungwa), Kenya_(Marsbit, Amboseli & Masai Mara), Ghana (Dusie, Jang, Gushiegu and Kabampe), Botswana, Namibia.

Research Hypothesis

The following hypotheses guide this study:

- Null Hypothesis: There is no significant difference in genetic diversity between spotted hyena populations across different regions.
- Alternative Hypothesis: There is a significant difference in genetic diversity between spotted hyena populations across different regions.
- There will be significant genetic structuring between regions in the north-western vs the eastern and southern population. In-depth structure is also expected within regions.

2. Methods

2.1. Sample details

This study includes data from 12 previously published genomes (Westbury *et al.*, 2020) and 61 restriction site-associated DNA sequences sampled from seven countries in Africa (Figure 2). The 61 RAD-Seq samples were provided by Prof Rasmus Heller of the University of Copenhagen. The laboratory analyses of the sequences are detailed in the Westbury *et al.*, (2021); however, brief details of the molecular procedure will be given below. The sampled countries include Botswana (n = 1), Ghana (n = 9), Kenya (n = 30), Tanzania (n = 23), Zambia (n = 3), Namibia (n = 1) and Somalia (n = 1).

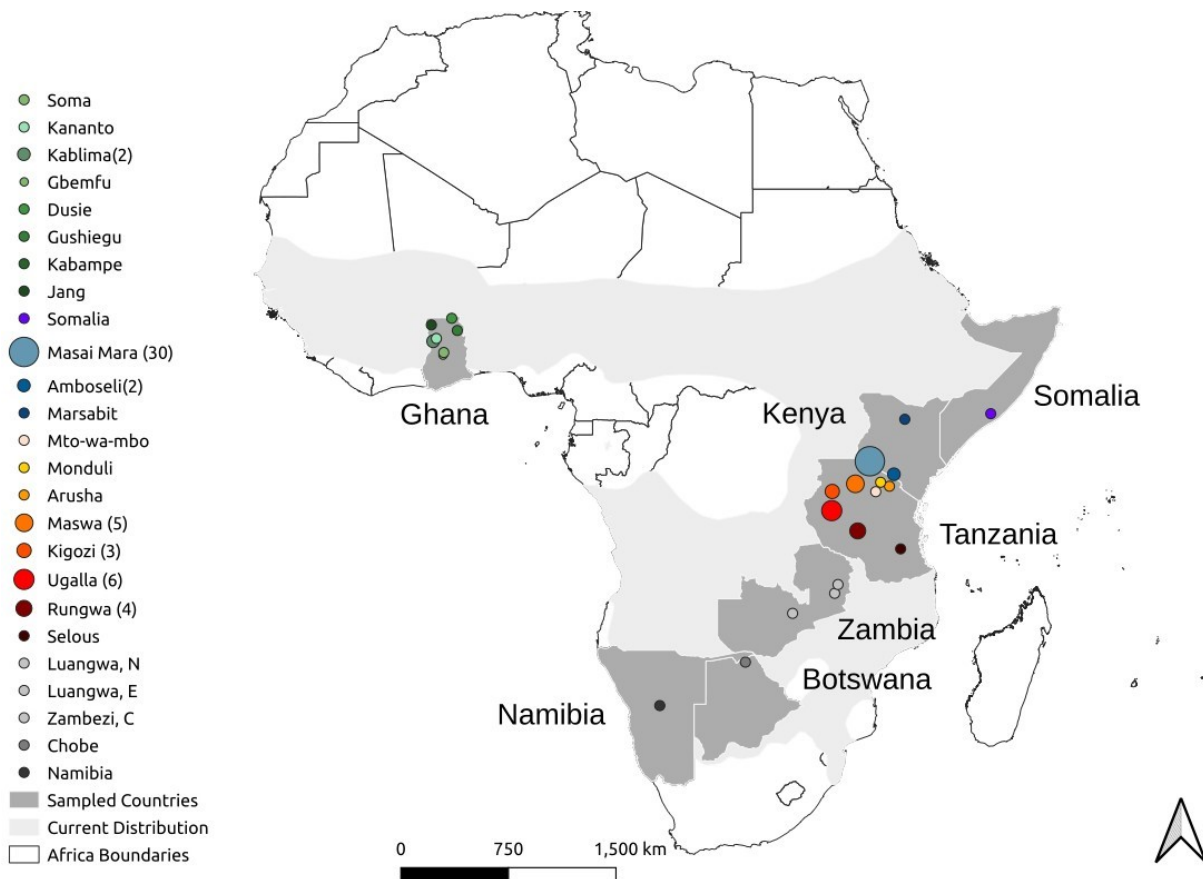


Figure 2: African map illustrating the current distribution of spotted hyena populations and sampled countries, furthermore there is specific regional information about where these spotted hyenas were sampled.

2.2. Reference and Outgroup

Reference genome assemblies are essential for any genomic study because DNA sequences are sequenced in a random order, therefore by using reference assemblies, sequences can be aligned to a unique location on the genome. Outgroups are mainly used in phylogenetic studies due to the assumption that ancestral states of a study group can be estimated using the character states of the outgroup, this approach is evaluated by the parsimony criterion.

There are only three assemblies of spotted hyena reference genomes available from online databases. The first was a DNA zoo sample https://dnazoo.s3.wasabisys.com/index.html?prefix=Crocuta_crocuta/ with a scaffold N50 length of 2.55 Gb that was sequenced in an unrecognised manner; therefore, it was not considered for this project. The second genome <https://ngdc.cncb.ac.cn/gwh/Assembly/17879/> shows assembly

was generated with a scaffold N50 length of 7.2 Mb, so it was not considered for the project. The last genome assembly was generated with a scaffold N50 length of 14 Mb [GenBank accession] <https://www.ncbi.nlm.nih.gov/sra/SRX6480028> (Shao *et al.*, 2022). The outgroup for this study was the striped hyena as the sister taxa for the spotted hyena. The two species are estimated to have diverged 8.6 Ma (Turner, 2008).

2.3. Data processing

The objective of this section is to process and transform data by alignment, filtering, quality control and variant calling.

2.3.1. Mapping

The individual fasta.qz data are simply files that consist of reads, which are mainly fragments in a library that are sequenced. The sequences are aligned with a reference genome assembly so that genetic variants can be identified, such as SNPs, and therefore I can study their distribution across the genome to further understand how they affect biological processes in relation to genetic diversity, structure and gene flow.

The spotted hyena data were mapped to the reference assembly genome using BWA-MEM v0.7.17 (Li *et al.*, 2013) and Samtools v1.13 (Li *et al.*, 2009) was used to parse the mapped files. Samtools generated BAM files that are simply compressed binary files containing sequence reads that are mapped to the reference sequence and the position of the reference sequence.

The mapped files were each sorted and then the duplicates were removed using samtools “remove dup” function.

2.3.2. Quality of the mapped data

BAM (Binary Alignment Map) files also show the mapping quality of the aligned sequence and the number of mapped reads. In this section, the objective was to remove low-coverage individuals from further analyses because low-coverage data causes bias in genetic studies.

Samtools depth was used to calculate the average depth of each BAM file and the number of reads mapped per individual. Individuals that were less than 2x coverage were removed from the dataset due to low resolution (Han *et al.*, 2014). See Table 1.

2.3.3. Autosomal and sex-linked scaffold identification

This study required the use of nuclear scaffolds; therefore, before the sampled data was analysed, any sex-linked scaffolds were removed to avoid varying rates of evolution and inheritance (Wilson Sayres, 2018).

To identify any sex-linked chromosomes, the spotted hyena reference chromosomes were compared with those from the cat X chromosomes and human Y chromosomes via Satsuma (Grabherr *et al.*, 2010; Liu *et al.*, 2018), which is a sequence alignment program that finds sequence matches through cross-correlation (Grabherr *et al.*, 2010). The hyena scaffolds that aligned with the cat X and human Y chromosomes were identified and extracted. Scaffolds that were less than 100kb were also extracted. This ensures that any mitochondrial DNA is removed, as any mitochondrial scaffolds are usually less than 16kb. The scaffolds that only contain autosomal scaffolds were then converted to a BED (Browser Extensible Data) file.

2.3.5. Relatedness

To avoid uncertainties with the data, ANGSD was used to calculate genotype likelihoods instead of directly calling genotypes because of the low-coverage dataset

ANGSD (Korneliussen *et al.*, 2014) was used to calculate the genotype likelihood including the following flags: the genotype likelihood were estimates using the GATK method (-GL 2), the dataset was filtered in the following manner: excluded reads with a mapping quality that is less than 25 (-

minmapQ 25), excluded bases with base quality less than 25 (-minQ 25), a minimum minor allele frequency of 0.05 or greater (-minmaf 0.05), exclude SNPs if the p-value is less than $1e-6$ (-SNP_pval $1e-6$), infer minor and major alleles from the genotype likelihoods (-doMajorMinor 1), only include sites if at least 90% of the individuals are covered (-minInd 50), exclude sex scaffolds and scaffolds shorter than 100kb (-rf), call allele frequencies based on a fixed major and an unknown minor allele (-doMaf 2). The maf output file was used to get the allele frequencies per site with the 'zcat' command flag.

ngsRelate (Korneliussen & Moltke, 2015) was used to identify closely related individuals from the dataset to avoid skewed data. Individuals that showed a high degree of relatedness (0.4) were removed from the dataset. When two individuals indicated a high degree of relatedness, the individual with the lower coverage between the two was removed from the dataset.

2.3.6. SNP Calling

In this study, I utilised both genotype likelihood and genotype calling for analyses such as the phylogenetic analysis, as well as gene flow, which enhances the accuracy of the analyses by comparing results from different methodologies.

The BAM files were converted into VCF (Variant Call Format) using bcftools (Danecek *et al.*, 2021) using bcftools view to identify the number of SNPs in the dataset. Filtering is required for these analyses because of sequencing errors and abnormalities that arise during mutations, such as multiallelic loci, and when comparing population studies, unique alleles may influence the results of the study.

The autosomal VCF files were then filtered, using the threshold --maf 0.05. This filters out sites that do not have a minor allele frequency (maf) equal to or greater than 0.05. The options-min-alleles 2 and -max-alleles 2 were used to filter out sites that were not biallelic, and this helps with producing accurate heterozygosity estimates. The option-- max-missing 0.1 was utilised to exclude sites that had more than 10% of missing data; anything more than 10% of missing data will produce significant bias estimates.

The option-- minQ 25 was used to filter out sites that had coverage of less than 25; these commands were comparable to those used for the genotype likelihood to maintain a standardised approach. Finally, the command --remove-indels was utilised to remove indels, which are variants that change the length of the reference allele. For the final SNP numbers before and after filtering see Table 2.

2.4. Diversity

Null: There will be no significant difference in genetic diversity between populations.

The main objective of this section is to assess the heterozygosity of spotted hyena populations in Africa compared to other large mammals that are sharing the same habitat with the spotted hyenas. Heterozygosity is ideal measurement of genetic diversity since low genetic diversity is related to inbreeding and inbreeding leads to reduced fitness, it can be assumed that heterozygosity and fitness at population levels are correlated (Reed & Frankham, 2003).

2.4.1. Heterozygosity

The heterozygosity of all the individuals was estimated using ANGSD with the following flags: the Samtools method was used to calculate genotype likelihood (-GL 1), the reference sequence (-anc) was used along with the flag (-fold 1) to fold the site frequency spectrum because the ancestral sequence was not available, calculated the site allele frequencies (-doSaf 1) and the reference sequence (-ref). we did not restrict our analyses to filtered sites, to be able to compare our estimates to published genome-wide heterozygosity values of other felids

The output files (the site allele frequencies) were used to calculate the site frequency spectrum using realSFS with the following flags: number of sites to use per iteration (-nSites 200000), tolerance for likelihood values (-tole 1e-8). The output file contained all homozygous positions and heterozygous positions in the dataset. The number of heterozygous positions was divided by the sum of the homozygous positions and heterozygous positions to get a heterozygosity estimate of the individual.

2.5. Population structure analyses

Hypothesis 2: There will be significant genetic structuring between regions

In this section, the population structure of the sampled spotted hyena populations is analysed using two different methods, firstly the model-free methods (PCA, and genetic distance) and the model-based method (model-based phylogeny and admixture)

2.5.1. Principal Component Analysis

The following flags were used to construct a beagle file for the PCA and admixture analysis: (-GL 2) (-minmapQ 25) (-minQ 25) (-minmaf 0.05) (-SNP_pval 1e-6) (-doMajorMinor 1) (-minInd 50) (-rf) and specified the output file as a beagle file (-doGlf 2).

For the PCA, PCangsd was used with default parameters, which required the input file (-beagle) and the output directory (-o), to estimate the individual allele frequencies and construct the frequencies into a covariant format file. The covariance format file was in R Studio, where it was used to visualize the PCA plot..

2.5.2. Admixture Analysis

The objective of the admixture analysis is to assess and identify the number of ancestral populations within the spotted hyena dataset.

The same beagle file was then used to estimate the admixture proportions between the individuals using the default parameters, the input file (-likes), number of ancestral populations (-K) and (-o) output file, NgsAdmix was used to model K2 to K10 ancestral populations, with each having 10 runs. The output file was used in an online software, Clumpak <https://clumpak.tau.ac.il/>, to identify the best K and visualise the admixture plots.

2.5.3. Phylogenetic analysis

2.5.2.1. Genetic distance based phylogenetic tree

A consensus base call approach (-IBS) was used on ANGSD with the same command flags from the PCA and Admixture genotype likelihood to create a distance matrix (-makeMatrix). The distance matrix was used to reconstruct a neighbour joining tree in nexus file format using PHYLIP neighbour.

2.5.2.2. Model-based phylogenetic tree

To create a model-based phylogeny using IQTREE to compare the phylogenies created based on genotype likelihood vs genotype calling, VCF file comprising of all the sampled hyenas was used instead of genotype likelihood because IQTREE required variant call formatted files as input to run the analysis and genotype calling is recommended for more accurate analysis especially when dating a phylogenetic tree. The VCF file including the outgroup were converted into a PHYLIP format using the script vcf2phylip whereby the SNP genotypes are used to produce a PHYLIP matrix for phylogenetic inferences. Once converted the output file has a mini.phy suffix which can be used by the program IQTREE to reconstruct the phylogeny (Nguyen *et al.*, 2015).

iqtree2 infers phylogenetic trees by maximum likelihood (ML) for the autosomal scaffolds. iqtree2 automatically chooses the best model based on the lowest Bayesian Information Criterion (BIC)

scores, but in this study, the GTR model was used to infer the phylogeny, because of extensive and long computational hours finding the best K was challenging; hence, the GTR model was selected based on other studies, which produced favourable results. The phylogenies were constructed using the following command flags: DNA sequences (-st DNA), specify the model's name to use during the analysis (-m GTR+R6), bootstrap support was 1000 (-b 1000). The bootstrap support mainly checks how many times the same branch appeared generating the phylogenetic from a 1000 runs (Minh *et al.*, 2019).

2.6. Gene flow

The main aim of this section is to analyse gene flow between populations, which allows us to infer the amount of gene material being shared between populations and also measure the connectivity between populations.

The D and F4-ratio statistics are methods that are used to calculate the correlations of allele frequencies across populations. The D-statistic will be used to estimate the introgression within the populations of the spotted hyena, under the assumptions that the species share a substantial amount of genetic variation due to common ancestry, incomplete lineage sorting, and recurrent and back mutants at the site are negligible.

D and f4-ratio statistics are used to assess the relative sharing of alleles among four populations or taxa, based on a rooted tree (((P1, P2), P3), O). In this case, O represents the ancestral allele, which is labelled A, while the derived allele is labelled B. This gives rise to the ABBA-BABA patterns whereby BBAA shows that P1 and P2 share the derived allele, ABBA means that P2 and P3 share the derived allele, and BABA means that P1 and P3 share the derived allele (Malinsky *et al.*, 2021).

Using the D-Suite package (Malinsky *et al.*, 2021), a command called Dtrios was used to infer population relations based on prior data such as a tree. This method includes fitting a phylogenetic tree model to a quartet of populations to provide a test for the history of admixture. For this study, the populations were defined based on the phylogenetic tree and population structure analyses from method section 2.5.

3. Results

3.1. Coverage

The detailed table (Table 1) shows all the samples by origin, park of origin, codename, coverage and number of reads. In terms of coverage the shaded cells represent samples that had a coverage of less than 2x and were removed from further analyses (see section 2). The only sample from Namibia had the highest coverage of 47,98 and highest number of mapped reads, while sample 7302 from Ghana had the lowest coverage from the dataset.

Table 1: All detailed information on the sampled individuals before analysing the data, with the Origin, Park of Origin, Codename, Coverage and Number of reads mapped. Shaded cells indicate samples excluded due to low coverage (<2x).

Origin	Park of origin	Codename	Coverage(x)	Number of reads mapping
Botswana	Chobe W-Kachikau	1979	6.73759	3826917
Ghana	Kablima-N/R	7302	1.25156	44584
Ghana	Soma	8126	1.25793	53117
Ghana	Kablima-N/R	7301	1.26355	48993
Ghana	Kananto-N/R	7489	1.26404	51965
Ghana	Gbemfu-N/R	7303	1.3058	67375
Ghana	Jang-N/R	7220	6.47846	1655088
Ghana	Kabampe-N/R	7490	12.9236	5789795
Ghana	Gushiegu, N/R	3858	21.18	21349084
Ghana	Dusie-UW/R	7221	47.9877	783317045
Kenya	Amboseli	827	1.52009	217602
Kenya	Marsabit	7708	3.58701	642619
Kenya	Masai Mara	799	10.1705	4361685
Kenya	Masai Mara	810	10.3095	6256490
Kenya	Masai Mara	796	11.6644	12036338

Kenya	Masai Mara	818	11.8012	10022988
Kenya	Masai Mara	812	12.1323	11823376
Kenya	Masai Mara	804	13.0841	12892373
Kenya	Masai Mara	811	13.7994	11452765
Kenya	Masai Mara	823	14.013	14324934
Kenya	Masai Mara	813	14.4897	10904948
Kenya	Masai Mara	817	14.5288	11491872
Kenya	Masai Mara	815	16.6035	12061098
Kenya	Masai Mara	806	16.8395	15467216
Kenya	Masai Mara	808	16.9235	16709699
Kenya	Masai Mara	809	17.4596	14651367
Kenya	Masai Mara	794	17.7645	16066164
Kenya	Masai Mara	819	17.7804	15083396
Kenya	Amboseli	826	18.1701	15934395
Kenya	Masai Mara	801	18.2127	16596772
Kenya	Masai Mara	820	18.5467	15848328
Kenya	Masai Mara	800	18.8263	16203282
Kenya	Masai Mara	797	19.1618	15871139
Kenya	Masai Mara	798	19.2855	16902201
Kenya	Masai Mara	821	19.8984	16014318
Kenya	Masai Mara	824	20.0011	16194351
Kenya	Masai Mara	807	20.3095	18708500
Kenya	Masai Mara	825	20.7245	16842633
Kenya	Masai Mara	802	20.9007	20461277
Kenya	Masai Mara	805	23.2187	21077061
Kenya	Masai Mara	803	24.1282	21368530
Kenya	Masai Mara	793	24.5901	31846653
Kenya	Masai Mara	795	24.7908	20989026
Namibia(wgs)			55.335	918733795

ZooSample		375	6.64483	110643981
Outgroup	Burko	3367	6.24905	2559429
Tanzania	Kigosi	8646	1.30634	102073
Tanzania	Mto-Wa-Mbu	3632	1.32266	110547
Tanzania	Ugalla	3631	1.36308	119256
Tanzania	Maswa	3633	1.45395	174218
Tanzania	Rungwa	3351	2.02589	445184
Tanzania	Ugalla	8645	2.96799	997327
Tanzania	Ugalla	5190	2.98453	1022290
Tanzania	Ugalla	3353	4.26506	1580621
Tanzania	Arusha	8416	4.34969	1031284
Tanzania	Kigosi	3352	4.42302	2160237
Tanzania	Kigozi	8643	5.67514	1758668
Tanzania	Maswa	3354	6.13501	3514462
Tanzania	Maswa	8642	6.33783	1579306
Tanzania	Ugalla	8644	6.3857	2287025
Tanzania	Selous	3630	6.46035	3463395
Tanzania	Ikiri-Rungwa	4036	8.15883	5872470
Tanzania	Maswa	4039	8.34858	6988314
Tanzania	Ikiri-Rungwa	4035	8.84724	5119119
Tanzania	Ugalla	5695	10.0068	9059390
Tanzania	Maswa	5696	11.696	11687260
Tanzania	Ikiri-Rungwa	4038	11.7959	9798258
Tanzania	Monduli	4037	14.6967	11227191
Zambia	Luangwa, E.	6377	1.28345	94752
Zambia	Zambezi, C.	6378	1.28867	102680
Zambia	Luangwa N.	2518	10.5663	7803526
ZooSample		2071	15.6426	257605923
Somalia			58.2613	13240090

3.2. SNPs data

The following table (Table 2) represents the results of filtering the dataset using the parameters set (see section 2.3.5 & 2.3.6). The autosomal markers were filtered using the same filtering methods but different software, ANGSD and Vcftools, to generate genotype likelihood sites and SNP the results yielded 63082 for genotype likelihood while 380721 for SNPs were generated for the analyses.

Table 2: The number of SNPs before and after filtering

	Unfiltered	Filtered
Autosomal SNPs	15255201 Sites	38072
Autosomal genotype likelihood sites		63082

3.3. Relatedness

The following table (Table 3) represents all the samples that have a relatedness of $r_{ab} > 0.4$. All the individuals from Ghana had a high degree of relatedness but were not removed from the dataset. A few samples from Masai Mara also showed a high degree of relatedness. Individuals from Masai Mara with the code names 797, 800, 803, 804, & 824 were removed from further analyses.

Table 3 : illustrates the results from ngsRelate with only the significant relatedness shown for more details about these results (see APPENTIX E). Only individuals with a relatedness (r_{ab}) of greater than 0.4 were selected for this table.

Ida	idb	r_{ab}	Fa	Fb
Ghana_Gushiegu 3858	Ghana_Dusie 7221	0.897922	0.463082	0.321086
Ghana_Jang 7220	Ghana_Dusie 7221	0.518375	0.342712	0.508064
Ghana_Dusie 7221	Ghana_Kabampe 7490	0.474868	0.524286	0.393745
Kenya_Masai Mara 794	Kenya_Masai Mara 797	0.429568	0.013307	0.014171
Kenya_Masai Mara 797	Kenya_Masai Mara 804	0.413105	0.013559	0.001306
Kenya_Masai Mara 797	Kenya_Masai Mara 805	0.424564	0.014327	0.012328
Kenya_Masai Mara 800	Kenya_Masai Mara 803	0.436957	0.062323	0.024808
Kenya_Masai Mara 800	Kenya_Masai Mara 815	0.437732	0.058263	0.06297

Kenya_Masai Mara 800	Kenya_Masai Mara 823	0.409632	0.057416	0
Kenya_Masai Mara 801	Kenya_Masai Mara 824	0.413306	0.009521	0.027735
Kenya_Masai Mara 803	Kenya_Masai Mara 815	0.435984	0.024076	0.063505
Kenya_Masai Mara 803	Kenya_Masai Mara 823	0.414605	0.019768	0
Kenya_Masai Mara 804	Kenya_Masai Mara 805	0.404188	9.00E-06	0.008151
Kenya_Masai Mara 808	Kenya_Masai Mara 825	0.426007	0.028278	0.01981
Kenya_Masai Mara 818	Kenya_Masai Mara 824	0.400761	0.014896	0.026704

3.4. Heterozygosity

The heterozygosity data indicate a range of heterozygosities across different regions. Somalia has the highest heterozygosity value of 0.0013, followed by Rungwa at 0.0011 also indicating a high level of heterozygosity. Regions such as Luangwa (0.0010), Amboseli (0.0010) and Monduli (0.0010) show moderate to high heterozygosity. While regions such as Masai Mara (0.0009), Chobe (0.0009), Jang (0.0009), Ugalla (0.0009), Kigozi (0.0008) and Maswa (0.0008) indicate average spotted hyena heterozygosity according to the dataset. Finally, Arusha (0.0006), Kabampe (0.0006), Gushiegu (0.0005) and Dusie (0.0003) have the lowest heterozygosity values with the dataset.

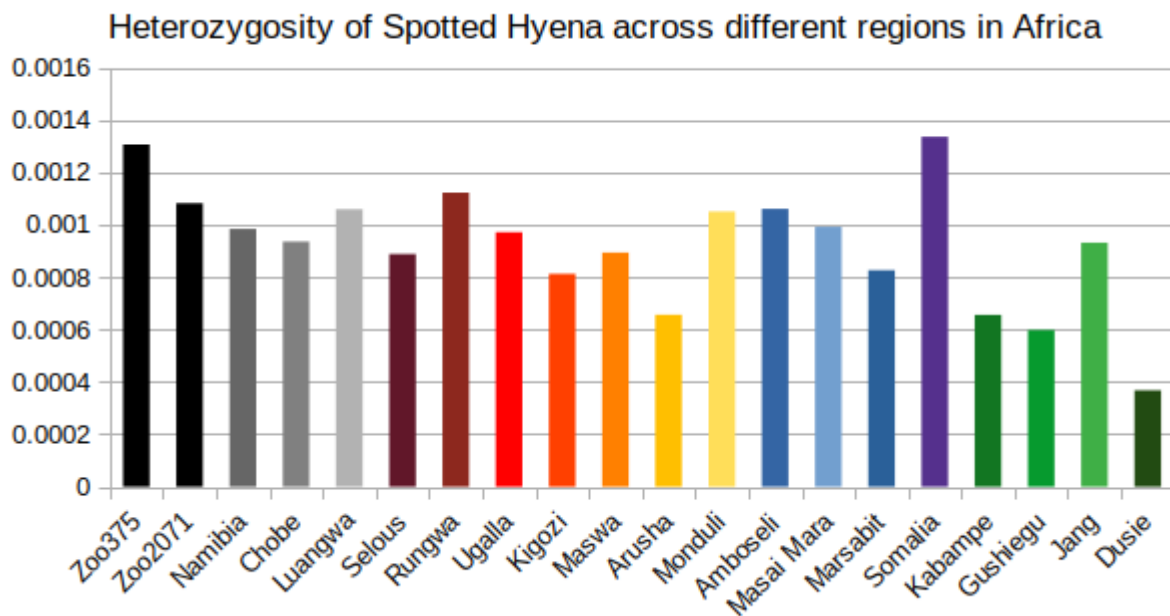


Figure 3: Heterozygosity of spotted hyena populations in sampled areas across Africa.

Samples Zoo_375, Zoo_2071, Botswana, Somalia & Namibia are WGS with a mean heterozygosity of 0.0011, while the rest of the RAD-seq have a mean heterozygosity of 0.00094. The rest of the sampled individuals are RAD-seq samples. Colours are specifically assigned to these regions in all further analyses.

3.5. Population Structure

3.5.1. Principal Component Analysis

The following figure (Figure 4) shows a model-free structuring of spotted hyena individuals based on their allele frequencies using PCA analysis. PC1 shows the separation of a West African population (Ghana) along with one Zoo sample, from the southern (Namibia, Botswana & Zambia) and eastern (Tanzania and Kenya) populations. PC2 showed a clear gradient from south to north, separating southern and eastern populations from the bottom centre to the top right corner. Overall, there was underlying phylogeographic structure within East Africa, with southern Tanzanian (Selous and some Rungwa) clustering with southern Africa, and separated from western Tanzania (Ugalla, Kigosi and Rungwa), which in turn was separated from the eastern Tanzania and southern Kenya (Maswa, Arusha, Monduli, and Amboseli). The Masai Mara population was further separated, with an internal structure of its own. The northern Kenya (Marsabit) and Somali individuals were intermediately separated on both PC1 and PC2. PC1 vs PC3 & PC2 vs PC3 showed similar levels of structure between the populations (see APPENDIX A)

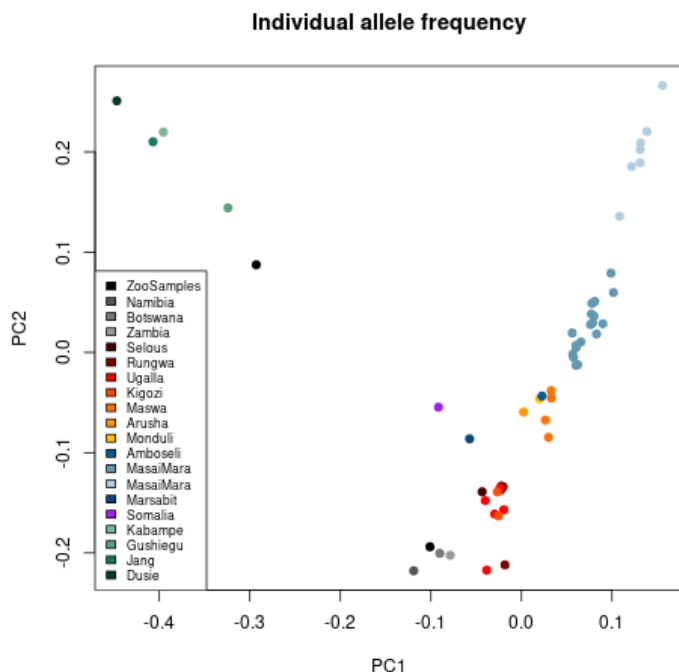


Figure 4: PC1 vs PC2 illustrating the individual allele frequency of spotted hyenas across sampled regions of Africa. * Both captive individuals are coloured in black.

3.5.2. Admixture proportions

The best K cluster was K3 which had the highest change in Delta K based on figure 5a below, while there was no change in Delka K between K4 and K9. Delta K is simply the calculated mean value of $L''(K)$ divided by the standard deviation of $L(K)$ as shown on figure 5a.

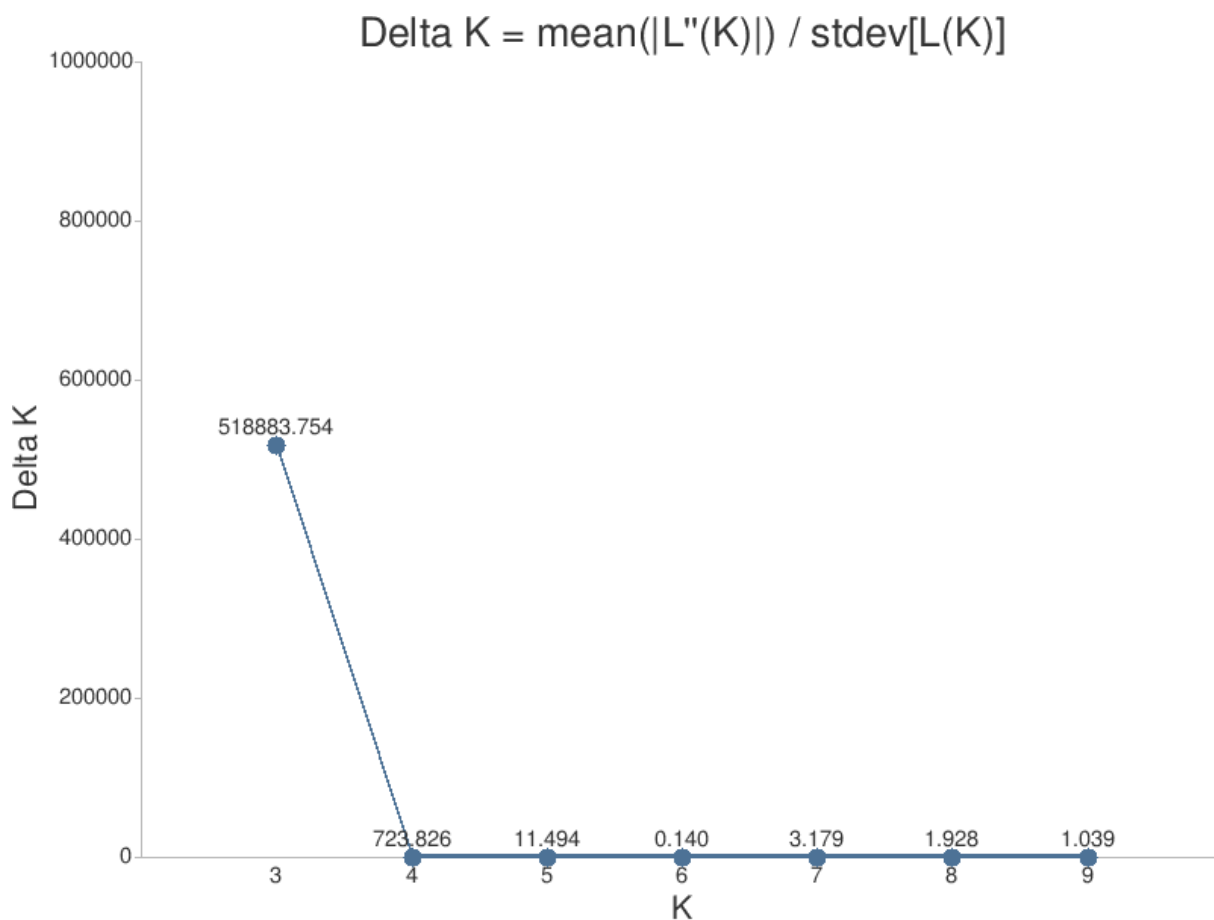


Figure 5a: Illustration of the Best K estimated using Clumpak whereby Delta K represents a statistic based on the rate of change in the log probability of Data between K3-K10.

The figure below (Figure 6b) shows the admixture proportions of K5 integrated into the geographical map of Africa to display the phylogeography of spotted hyena populations.

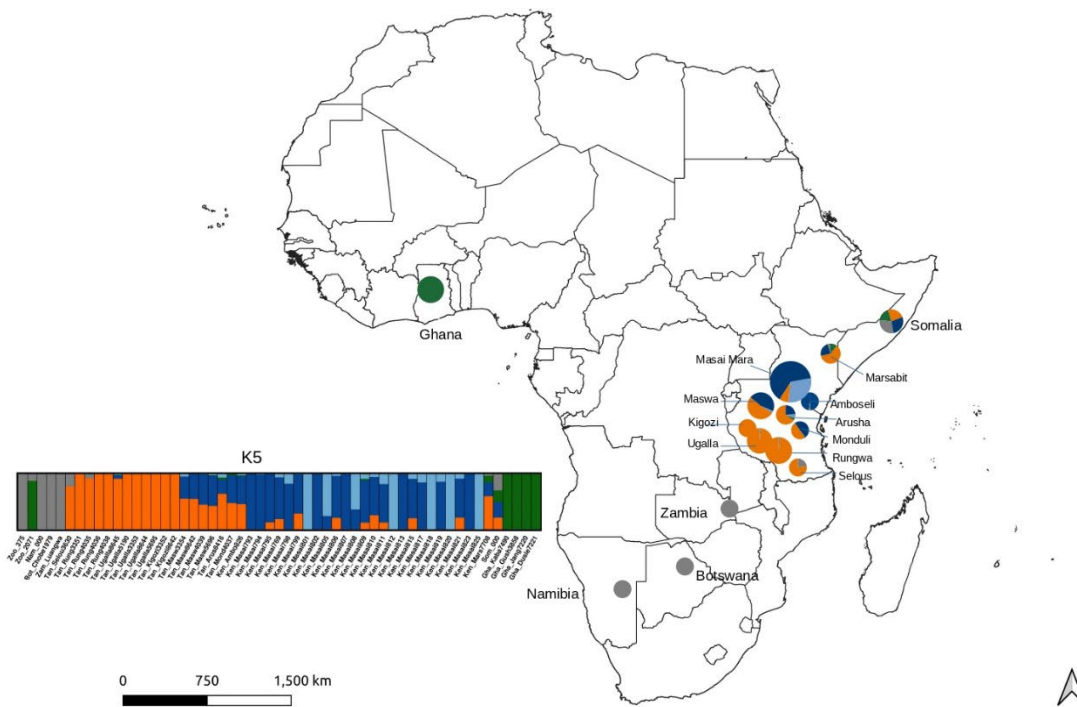


Figure 6b: Admixture proportions of K5 integrated onto the map of Africa using pie charts, the pie charts are resized based on the size or number of individuals within that population.

3.5.3. Phylogenetic Analysis

The phylogenetic analysis of spotted hyena populations across Africa indicates five main clades (Figure 7), with the first to diverge from the group being the western population (Ghana), followed by the north/east individuals (Marsabit and Somalia). The next branch to diverge from the main group was the southern populations (Zambia, Botswana, Namibia & Selous), while the next branch included all other Tanzanian and Kenyan samples. Within this group, western Tanzania (Ugalla, Rungwa, Kigozi) and one eastern Tanzania individual (Arusha) were ancestral and separated from the eastern Tanzania and Kenya, which consisted of Maswa, Monduli, Amboseli and Masai Mara. Zoo_2071 clustered with the West African populations, while Zoo_375 clustered with the southern African populations. Within some individuals (light blue) were more derived than others.

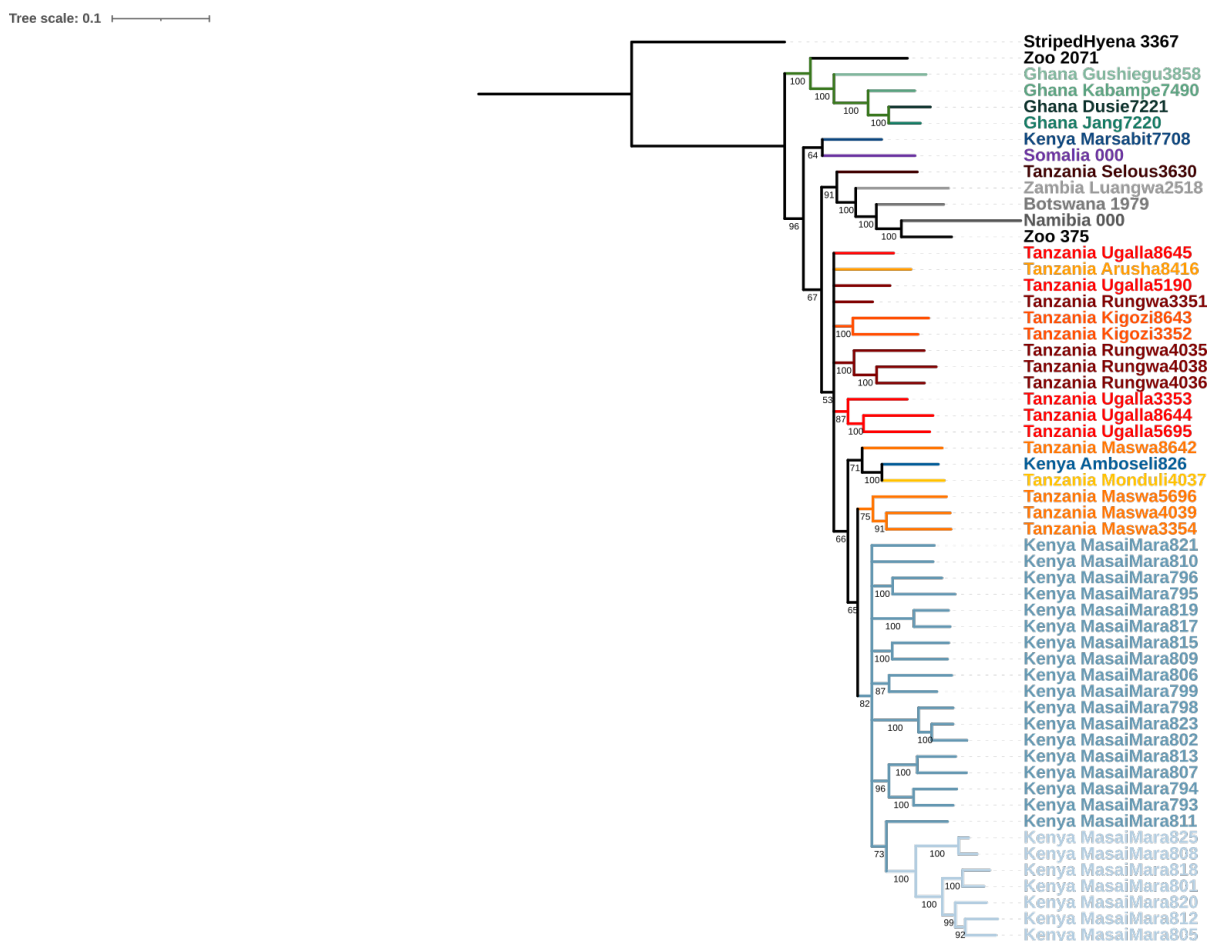


Figure 7: Phylogenetic tree of spotted hyena across sampled regions of Africa, branches with a bootstrap support of less than 50% collapsed. A comparable phylogenetic tree can be found in APPENDIX C that was based on genetic distance matrix of the spotted hyena using ANGSD.

3.6. Gene flow

Table 4 indicates the gene flow results from D-Suite using the following phylogenetic tree as a reference base in the following newick format: (Outgroup,(West,(NorthEast,(Southern,(westernTanzania,easternTanzania/Kenya)))); with the outgroup being the striped hyena, West (Dusie, Gushiegu, Kabampe & Jang), NorthEast (Marsabit & Somalia), eastern Tanzania Kenya (Arusha, Maswa, Moduli, Amboseli & Masai Mara), westernTanzania (Kigozi, Ugalla, Rungwa) and Southern (Namibia, Botswana, Zambia & Selous).

These results only show significant Z-scores based on the given phylogenetic tree; there is another table of D-stats, which was generated using ANGSD without a reference tree, see APPENDIX D, giving all possible combinations. Eastern Tanzania Kenya & NorthEast share the most significant amount of gene flow with a Z-score of 7.71871 in this whole dataset followed by Southern & westernTanzania with a Z-score of 6.16682. On the other hand, eastern Tanzania, Kenya and Northeast share the least amount of gene flow in this dataset with a Z-score of 1.255518, followed by eastern Tanzania, Kenya and western Tanzania with a Z-score of 1.9273, both of which were not significant (not more than 3).

Table 4: The gene flow results estimated using D-Suite with a reference tree, another detailed d-stats results that do not include the reference tree is included in the APPENDIX. This table shows the gene flow between P1, P2 & P3, with the striped hyena as the outgroup. Significant values of Z-scores are also shown.

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
Southern	TanzaniaKenya	NorthEast	0.0418439	7.71871	1.17511E-14	0.0711948	1452.57	1534.17	1410.94
Southern	NorthEast	Western	0.0369022	5.1078	3.25924E-07	0.0972657	1462.36	1490.7	1384.59
Southern	WesternTanzania	NorthEast	0.0255141	5.58979	2.27345E-08	0.0425106	1466.09	1478.82	1405.23
TanzaniaKenya	NorthEast	Western	0.0081931	1.25518	0.209412	0.0229562	1520.84	1425.94	1402.76
WesternTanzania	TanzaniaKenya	NorthEast	0.0175748	4.36551	1.26827E-05	0.0298822	1521.96	1437.39	1387.74
WesternTanzania	NorthEast	Western	0.0198011	2.72316	0.00646611	0.0536743	1482.95	1437.7	1381.87
Southern	TanzaniaKenya	Western	0.0291339	6.16682	6.96775E-10	0.0768578	1521.77	1485.76	1401.64
TanzaniaKenya	WesternTanzania	Southern	0.0065833	1.9273	0.0539419	0.0303507	1566.23	1417.91	1399.36
Southern	WesternTanzania	Western	0.0178299	4.15525	3.24937E-05	0.0461736	1530.62	1442.47	1391.94
WesternTanzania	TanzaniaKenya	Western	0.012132	3.30233	0.000958865	0.0321798	1603.79	1400.91	1367.32

In Figure 8 almost all the populations share genetic material with each other with both westernTanzania and Tanzania Kenya being the hub of gene flow as these regions share genetic materials with all the sampled regions while the northeast only shows significant gene flow with the westernTanzania and Tanzania Kenya. The southern regions only share gene flow with the western Tanzania, while the western populations share gene flow with both the western Tanzania and Tanzania, Kenya.

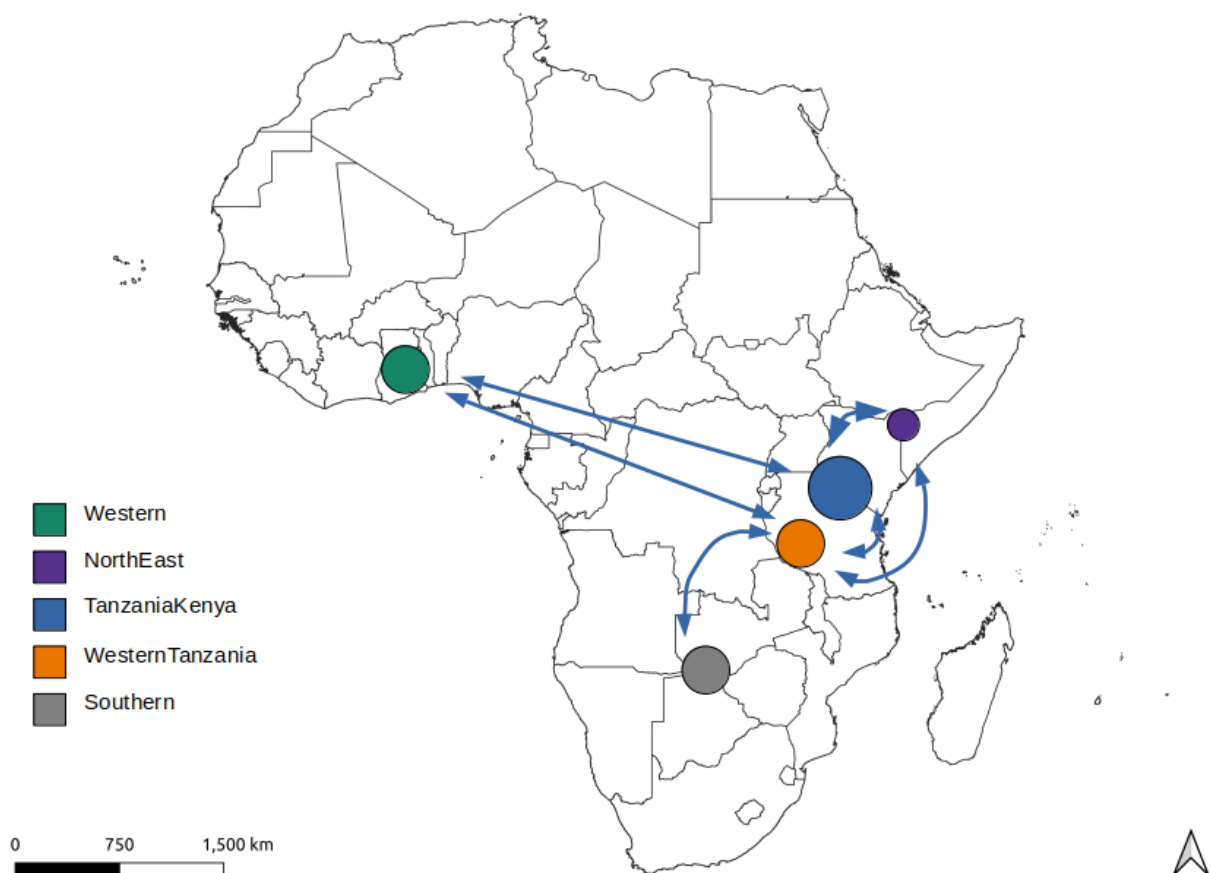


Figure 8: This figure illustrates the table (Table 4) but with the African map integrated to indicate the flow of genetic materials between populations. The arrows show the gene flow into and out of regions.

4. Discussion

The data processing steps began with data from restriction site-associated and whole-genome sequencing libraries. This data was collected from 72 individual spotted hyenas across Africa, 12 individuals were removed due to low coverage (See Table 1), and 5 individuals were removed due to high relatedness of greater than 0.4, as these individuals could be the second or third cousins. Therefore, the dataset was reduced from 72 to 55 sequences, mainly to avoid bias during population inferences. Another challenge encountered during the research was the sex determination of the spotted hyena, but methodologies that infer the sex of the individual from the sequences were mainly based on whole genome sequences and because much of our dataset consisted of restriction site-associated sequences, this chapter was left unsolved (see section 2.3.3).

Using synteny, I identified the autosomal scaffolds (392 autosomal scaffolds) from our references and then used those to separate our autosomal dataset from the mapped sequences while also removing any sites with less than 100Kb (see 2.3.3). Using the same filtering criteria, I generated genotype likelihood sites (63082 sites) and SNPs (38072 SNPs) (see section 3.2). Generating both genotype and SNPs was a way to compare different analyses, such as the genetic distance phylogeny and the model-based phylogeny. This was crucial in deciding the best tree for this study (see figure 7; APPENDIX C). In this study, I only managed to extract autosomal markers and X-chromosome markers, but was unable to extract mtDNA. This was mainly due to the nature of restriction site-associated sequences because of small fragments from the whole genome. From the X-chromosome, I identified 317 sites, but they were less informative (see APPENDIX D).

4.1. Genetic diversity

Figure 3, shows the heterozygosity of spotted hyena populations, Somalia (0.0013) has the highest heterozygosity value in the dataset, indicating a relatively high level of genetic diversity, similar Rungwa (0.0011), Luangwa (0.0010), Amboseli (0.0010) and Monduli (0.0010) also indicates a high

level of heterozygosity which is simply a sign of diversified gene pool within these regions. This suggests that the spotted hyenas in these regions have a robust gene pool, which is much more correlated with those populations with high admixture (Figures 5 and 6a). Such a diverse gene pool is an advantage in wild populations in terms of adaptability to environmental changes, disease resistance, and evolutionary potential.

Regions such as Dusie (0.00036), Arusha (0.00065), Kabampe (0.00065) and Gushiegu (0.00059) are the regions with the lowest heterozygosity. Most notably, the West African regions have very low heterozygosity. Low genetic diversity in wild populations is mainly a sign of concern from a conservation perspective, and it is a sign of a small population size and limited gene flow; both these factors increase the risk of inbreeding and reduced fitness. Although most of central Africa was not sampled in this dataset, the Ghana population is experiencing genetic deficiency, which could be due to historical bottlenecks within that region. Lack of gene flow from the eastern to western regions could also explain this low genetic diversity in the west, even though D-stat results show (figure 8) gene flow from and to the western region, it is definitely not enough based on the genetic diversity. Additionally, the average heterozygosity of spotted hyenas in West Africa is 0.00063, while the heterozygosity of spotted hyenas in Africa is 0.0009. Furthermore, it needs to be stated that the Ghanaian populations were all closely related to each other, hence there is a possibility that the data is skewed by high relatedness. However, it is also worth mentioning that the individuals were sampled in different regions within Ghana, not in the same area.

In comparison to African large carnivores such as the African lion (Smitz *et al.*, 2018) and leopards (Pečnerová *et al.*, 2021) that share the same sub-Saharan range as the spotted hyena and are well known to be competitors of the spotted hyena, reveal varying heterozygosities. Both these species have higher heterozygosity compared to the spotted hyena, with the African lion heterozygosity ranging from 0.00074 to 0.00148, while the African leopards have much higher heterozygosity, ranging from 0.0017 to 0.0036 (Pečnerová *et al.*, 2021). Although African lions also reveal similar patterns of genetic diversity with the spotted hyena. The genetic diversity of both species is low in the west-central lineage compared to east southern lineage. (Smitz *et al.*, 2018).

The African leopard show contrasting data with the west lineage indicating the highest heterozygosity values at around 0.036 while the East-Southern lineage also indicated relatively consistent levels of heterozygosity ranging from 0.0017 to 0.0020 (Pečnerová *et al.*, 2021), while the spotted hyena and lion seem to have relatively low genetic diversity in the west-central regions the leopards have the highest genetic diversity.

Relative to the other felids, the genetic diversity of spotted very moderate to low especially considering that the spotted hyena population is greater than that of the African lion or leopard but what is more concerning is that the western region of the spotted hyena diversity is ranging at 0.0003 to 0.009, hence it is more comparable to endangered species such as the cheetah (*A. j. hecki* range: 0.00020 to 0.00050; *A. j. soemmeringii* range: 0.00029 - 0.00050, *A. j. jubatus* range: 0.00036 - 0.00050; *A. j. raineyi* range: 0.00044 - 0.00048 and *A. j. venaticus* a mean of 0.0002 range: 0.00020 - 0.00040) (Prost *et al.*, 2020) and the brown hyena with heterozygosity well below 0.003 (Westbury *et al.*, 2018).

4.2. Population Structure

On the continental scale, the population structure of spotted hyena populations reveals three groups in the Southern, Eastern and Western populations. The southern population indicates a higher level of admixture with the eastern populations meaning that there is some form of gene flow between these two regions, similarly the eastern population is also largely admixed with the southern population, which can lead to high levels of genetic diversity between these two regions and heterozygosity of spotted hyena populations also support this narrative as the southern and eastern regions share similar levels of heterozygosity.

In contrast the western population is more genetically isolated, showing no signatures of admixtures with either the southern or western populations. Such a lack of gene flow between populations could be a sign of geographically isolated populations, which may lead to inbreeding or a more distinct identity within the western spotted hyena population. D-stats suggest that there is gene flow into and out of the western region, although it is not strong enough to disrupt this strong genetic structuring. However, no intervening populations were sampled especially in the central regions of Africa and also only one country was sampled in the western regions while all the sampled individuals showed a high degree of relatedness which can also be attributed to inbreeding or lack of inclusive samples but ultimately the data suggest that the western population is isolated from the east and southern populations.

Such population structure between large African mammals has been recorded before in other species such as the lion with the population structure separating the west-central vs east-southern populations of African lions (Pečnerová *et al.*, 2021), this phylogeographic structuring was also recorded with the hartebeest (Raballah, 2010) indicating a high degree of differentiation between the North and Southern regions of Africa, with the north mainly encompassing the west and eastern region and south consisting of the east and southern regions. Roan antelope and Common warthog showed similar genetic structuring, indicating that these species experienced the same

environmental pressures and evolutionary events as the spotted hyena, especially since they are all widely distributed across the sub-Saharan African savannahs (Alpers *et al.*, 2010; Muwanika *et al.*, 2003). Such correlations are expected between these species possibly because of predator-prey relationships but geographically the North/South split that separated these species for a period of time can mostly be attributed to the increase in tropical forest during the pluvials across Central Africa. Such pluvial cycles occurred during the Pleistocene, with at least five full pluvial cycles recorded during that time.

On a regional scale, the Kenyan population was structured, and the admixture analyses and PCA indicated that there were two populations within the Masai Mara. This could relate to two different hyena clans in the Masai Mara, although detailed geographic locations and other metadata were not available. One of these Kenyan populations contributed ancestry to eastern Tanzanian populations, which are neighbouring the Kenyan border. The Marsabit and Somali individuals shared the same admixture proportions, and on the phylogenetic tree, they were the second cluster to diverge from the rest of the populations following the divergence of the western populations. This also indicates that populations in southern and northern Kenya have different population histories. Microsatellite markers support the notion that the Amboseli and Masai Mara populations experience different population histories (Watts *et al.*, 2011). Such that differences were explained to be caused by the different patterns of male dispersal between the two regions, with the Masai Mara having more male immigrants than the Amboseli. Although the present study shows genetic separation between the Masai Mara and Amboseli, this theory cannot be affirmed in this study, as there is only one sample from Amboseli, and the sex of our sampled individuals could not be determined using RAD-seq data.

Tanzania clearly shows three distinct clusters: one in the south (Selous), one in the north (Maswa, Arusha, and Monduli), and one in the west (Ugalla, Rungwa, and Kigozi). The northern cluster is primarily admixed with populations from Amboseli and Masai Mara, while the western cluster does not show admixture with any other population based on the analysis of admixture proportions (Figure 6a) and the phylogenetic tree (Figure 7). The southern Tanzania cluster, which consists solely of Selous, displays signatures of admixture with southern African populations from Zambia, Namibia, and Botswana.. These three clusters within Tanzania can also be seen with African (Pečnerová *et al.*, 2021) lions and the sable antelope (Matthee & Robinson, 1999). This separation is mainly due to the anthropogenic factors and geographical barriers, namely the Eastern Arch Mountains, with the west and southern population situated on either side of the mountains, hence reducing gene flow between these areas (Matthee & Robinson, 1999; Pečnerová *et al.*, 2021). Genetic diversity between these remains largely similar, which indicates that the differentiation between these clusters is not that severe.

Looking the Botswana, Namibia and Zambia there is not much to detect on the regional scale mainly because of lack of sampled areas hence the regional or internal population structure cannot be inferred but it is worth noting that there is internal population structure within the Zambian population as recorded by Martens De Vry, 2021 based on a SNP panel. Unfortunately, these 83 samples couldn't be integrated into this study. That study revealed a separation between the Liuwa plain and the Kafue and Luangwa valleys. That population structure can be explained by the Zambezi River separating Liuwa from the other regions of Zambia.

4.3. Gene flow

On a continental scale, gene flow is mainly prominent between the east and southern spotted hyena populations. The high levels of admixture proportions between the east and southern regions indicate a high frequency of dispersal events, thus encouraging the exchange of genetic material between populations within these regions. This is also shown in their similar levels of heterozygosity suggesting that the regions are well connected, this is also highlighted by the significant Z-scores recorded between these populations. On the other hand, the western populations show very limited gene flow with both the east and southern populations, thus being a cause for concern as limited gene flow suggest genetic isolation between populations. The significant Z-scores between the east vs western or south vs western indicate low levels of gene flow between these populations. This limited gene flow between the western and south-eastern populations could be attributed to natural barriers such as the tropical forests and anthropogenic factors that prohibit the movement of spotted hyena population. Again, this study lacks samples from central African but based on the African lion study which included central African samples the same level of separation was evident hence there is no doubt that there is limited connectivity between these regions. This limited connectivity to and from the western populations is not ideal as it encourages the risk of bottlenecks that will lead to reduced adaptability and fitness.

In East Africa, gene flow is particularly evident between the Kenyan and Tanzanian spotted hyena populations as shown from admixture and phylogenetic analyses suggesting frequent dispersal across the Kenya-Tanzania border, this is mainly supported by the Amboseli population displaying strong genetic ties to the Tanzanian counterparts as seen on the phylogenetic tree. The high Z-score of between Tanzania/Kenya reinforces this significant gene flow between these regions. Furthermore, the Marsabit population share significant genetic similarities with the Somalian populations, suggesting historical or ongoing gene flow between these regions; hence, it can be assumed that individuals are free to disperse between these areas with minimum restrictions.

These findings are supported by statistical analyses, including D-statistics, which reveal significant genetic differentiation between Western populations and other regions. For instance, comparisons between Southern and Western populations yield highly significant p-values, emphasising their genetic separation. Similarly, Z-scores for gene flow between Eastern and Western populations indicate minimal admixture, reinforcing the idea of isolation. The genetic data also suggests varying levels of historical gene flow across Africa, with some regions maintaining high connectivity while others exhibit strong signs of genetic structuring. These patterns align with observations in other large African mammals, where environmental pressures and geographical barriers have historically shaped population dynamics.

4.4. Conservation Implications

The genetic diversity of spotted hyena is relatively high compared to endangered species such as the wild dog and the cheetah but when compared to their main rivals such as the African lion and leopard, the spotted hyena genetic diversity is low to moderate especially considering the fact that the spotted hyenas have far much larger populations compared to the lion or leopards. Particularly the genetic diversity of spotted hyena in the western regions are close to the levels of endangered species which is a concern because low genetic diversity affects the adaptability and disease resistance of populations, there need to be a serious intervention in the western population. Threats like urbanization, agricultural expansion, and infrastructure development are already invading on the spotted hyena habitats.

Eastern Tanzania Kenya and western Tanzania Kenya are close to each other yet there is genetic structuring within these regions suggesting that conservation strategies like corridors and translocations are required for forstering gene flow between these regions to avoid further genetic differentiation such strategies could also assist reducing the high structure found in the western regions (Beier & Noss, 1998). The western regions should be prioritised for conservation efforts as there is evidence of low genetic diversity compared to the other regions suggesting that this region could be isolated from the rest of the population of spotted hyena, moving few male immigrants from the east-southern populations to the western would reduce the degree of structure and encourage genetic diversity in the western region.

There is a need for initiatives and action taken towards conservation strategies, the establishment of protected areas and corridors may enhance connectivity between fragmented populations, promoting gene flow and enhancing genetic diversity. Such strategies can mainly be helpful in Zambia, eastern Tanzania, Kenya, western Tanzania and most parts of the western regions. We also need educational programs that focus on raising awareness of the impactful nature of spotted

hyenas in our ecosystem and along with reducing the stigma associated with spotted hyenas to reduce illegal hunting that is mainly driven by the idea that spotted hyena can be used as traditional medicine and other cultural practices (Lindsey *et al.*, 2013).

More genetic research is required in relation to the spotted hyenas in Africa to further understand the genetic diversity levels, especially within central Africa and also South Africa, which was largely not sampled in this study. Also, being on constant monitoring of population dynamics, behaviour, and changing territory ranges can inform conservation strategies on identifying areas where management interventions are needed.

References

Alpers DL, van Vuuren BJ, Arctander P, Robinson TJ., 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, 13, 1771–1784.

- Beier, P. and Noss, R.F., 1998. Do habitat corridors provide connectivity?. *Conservation biology*, 12(6), pp.1241-1252.
- Boydston, E.E., Kapheim, K.M., Van Horn, R.C., Smale, L. and Holekamp, K.E., 2005. Sexually dimorphic patterns of space use throughout ontogeny in the spotted hyena (*Crocuta crocuta*). *Journal of Zoology*, 267(3), pp.271-281.
- Cariou, M., Duret, L. and Charlat, S., 2016. How and how much does RAD-seq bias genetic diversity estimates?. *BMC Evolutionary Biology*, 16(1), pp.1-8.
- Ceballos, G., Ehrlich, P.R. and Dirzo, R., 2017. Biological annihilation via the sixth mass extinction signalled by vertebrate population losses and declines. *Proceedings of the national academy of sciences*, 114(30), pp.E6089-E6096.
- Cozzi, G., Broekhuis, F., McNutt, J.W. & Schmid, B., 2013. Density and habitat use of lions and spotted hyenas in northern Botswana and the influence of survey and ecological variables on call-in survey estimation. *Biodivers. Conserv.* 22, 2937–2956.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H., 2021. Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2), p. giab008.
- De Blocq, A.D. 2014. Estimating spotted hyaena (*Crocuta crocuta*) population density using camera trap data in a spatially-explicit capture-recapture framework (Doctoral dissertation, University of Cape Town).
- Ellerman, J.B., Morrison-Scoot, T.C.S. and Hayman, E.W., 1953. Southern African mammals 1758 to 1951: a reclassification. *Southern African Mammals 1758 to 1951: a Reclassification*.
- Ellegren, H. and Galtier, N., (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), pp.422-433.
- Engh, A.L., Funk, S.M., Horn, R.C.V., Scribner, K.T., Bruford, M.W., Libants, S., Szykman, M., Smale, L. and Holekamp, K.E., 2002. Reproductive skew among males in a female-dominated mammalian society. *Behavioral Ecology*, 13(2), pp.193-200.
- Frankham, R., Briscoe, D.A. and Ballou, J.D., 2002. Introduction to conservation genetics. *Cambridge University Press*.
- Garner, A., Rachlow, J.L. and Hicks, J.F., 2005. Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology*, 19(4), pp.1215-1221.

- Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F. and Lindblad-Toh, K., 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, 26(9), pp.1145-1151.
- Han, E., Sinsheimer, J.S. and Novembre, J., 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular biology and evolution*, 31(3), pp.723-735.
- Hayssen, V. and Noonan, P., 2021. *Crocuta crocuta* (Carnivora: Hyaenidae). *Mammalian Species*, 53(1000), pp.1-22.
- Hofer, H. and East, M., 1995. Population dynamics, population size, and the commuting system of Serengeti spotted hyenas. *Serengeti II: dynamics, management, and conservation of an ecosystem*, 2, pp.332-363.
- Hofreiter, M. and Barnes, I., 2010. Diversity lost: Are all Holarctic large mammal species just relict populations? *Journal of Biology*, 9(3), pp.1-3.
- Hofreiter, M. and Sheng, G., 2021. Ancient mitochondrial genomes from Chinese cave hyenas provide insights into the evolutionary history of the genus *Crocuta*. *Proceedings of the Royal Society B*, 288(1943), p.20202934.
- Hohenlohe, P.A., Funk, W.C. and Rajora, O.P., 2021. Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), pp.62-82.
- Holekamp, K.E., Sakai, S.T. and Lundrigan, B.L., 2007. Social intelligence in the spotted Hyaena (*Crocuta crocuta*). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), pp.523-538.
- Holekamp, K.E. and Dloniak, S.M., 2010. Intraspecific variation in the behavioural ecology of a tropical carnivore, the spotted Hyaena. In *Advances in the Study of Behavior* (Vol. 42, pp. 189-229). Academic Press.
- Holekamp, K.E., Smith, J.E., Strelhoff, C.C., Van Horn, R.C. and Watts, H.E., 2012. Society, demography, and genetic structure in the spotted Hyaena. *Molecular Ecology*, 21(3), pp.613-632.
- Hu, J., Westbury, M.V., Yuan, J., Zhang, Z., Chen, S., Xiao, B., Hou, X., Ji, H., Lai, X., Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J. and Willerslev, E., (2004). Rise and fall of the Beringian steppe bison. *Science*, 306(5701), pp.1561-1565.

Kalmykov, N.P. and Kobylkin, D.V., 2017. November. The first find of the cave hyena (*Crocota crocuta spelea*) on the Oka plateau (Eastern Sayan). In *Doklady earth sciences* (Vol. 477, pp. 1247-1250). *Pleiades Publishing*.

Koepfli KP, Jenks SM, Eizirik E, Zahirpour T, Van Valkenburgh B, Wayne RK., 2006. Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Mol Phylogenet Evol.* 38(3):603–620.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, 15(1), pp.1-13.

Korneliussen, T.S. and Moltke, I., 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics*, 31(24), pp.4009-4011.

Kruuk, H., 1972. The spotted hyena : a study of predation and social behavior. University of Chicago Press

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Lindsey, P.A., Havemann, C.P., Lines, R., Palazy, L., Price, A.E., Retief, T.A., Rhebergen, T. and Van der Waal, C., 2013. Determinants of persistence and tolerance of carnivores on Namibian ranches: implications for conservation on Southern African private lands. *PLoS One*, 8(1), p.e52458.

Liu, D., Hunt, M., and Tsai, I.J., 2018. Inferring synteny between genome assemblies: a systematic evaluation. *BMC bioinformatics*, 19(1), pp.1-13.

MacPhee, R.D., Tikhonov, A.N., Mol, D. and Greenwood, A.D., 2005. Late Quaternary loss of genetic diversity in muskox (*Ovibos*). *BMC Evolutionary Biology*, 5(1), pp.1-13.

Mathee CA, Robinson TJ., 1999. Mitochondrial DNA population structure of roan and sable antelope: implications for the translocation and conservation of the species. *Molecular Ecology*, 8, 227–238.

Mertens De Vry, A., 2021. De novo SNP Discovery in the Spotted hyena (*Crocota crocuta*).

Minh, B.Q., Trifinopoulos, J., Schrempf, D., Schmidt, H.A. and Lanfear, R., 2019. IQ-TREE version 2.0: tutorials and Manual Phylogenomic software by maximum likelihood. URL [http://www. iqtree. org](http://www.iqtree.org).

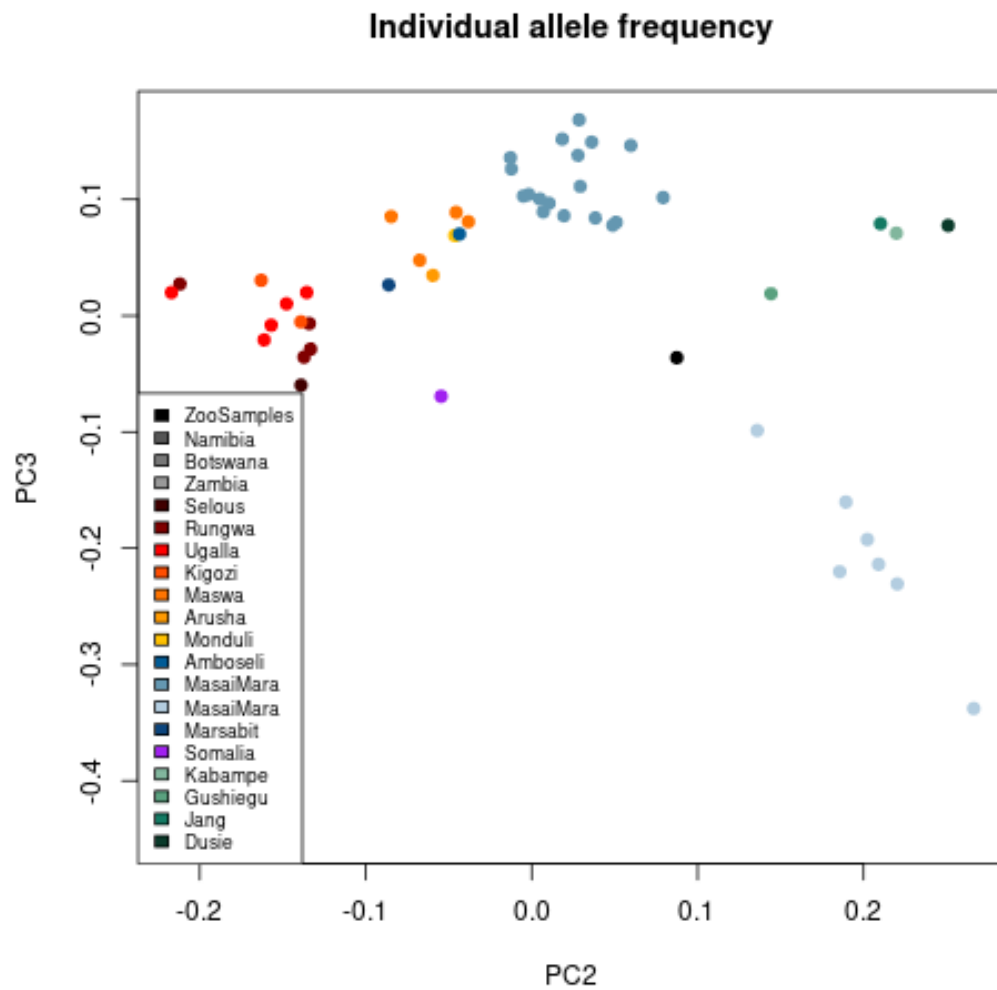
- Morin, P.A. & M. McCarthy., 2007. Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes*. 7(6): p. 937-946.
- Muwanika VB, Nyakaana S, Siegismund HR, Arctander P., 2003. Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, 91, 361–372.
- Ouma C, Roca A, Were T, Raballah E., 2010 Genetic structure of hartebeest populations straddling a transition zone between morphotypes. *Journal of Basic and Applied Scientific Research*, 1, 131–149.
- Palmqvist, P., Martínez-Navarro, B., Pérez-Claros, J.A., Torregrosa, V., Figueirido, B., Jiménez-Arenas, J.M., Espigares, M.P., Ros-Montoya, S. and De Renzi, M., 2011. The giant hyena *Pachycrocuta brevirostris*: modelling the bone-cracking behavior of an extinct carnivore. *Quaternary International*, 243(1), pp.61-79.
- Pečnerová, P., Garcia-Erill, G., Liu, X., Nursyifa, C., Waples, R.K., Santander, C.G., Quinn, L., Frandsen, P., Meisner, J., Stæger, F.F. and Rasmussen, M.S., 2021. High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Current Biology*, 31(9), pp.1862-1871.
- Prost, S., Machado, A.P., Zumbroich, J., Preier, L., Mahtani-Williams, S., Meissner, R., Guschanski, K., Brealey, J.C., Fernandes, C., Vercammen, P. and Hunter, L.T., 2020. Conservation Genomic Analyses of African and Asiatic Cheetahs (*Acinonyx jubatus*) across Their Current and Historical Species Range. *BioRxiv*, pp.2020-02.
- 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(12), pp.2435-2443.
- Schlötterer, C., 2004. The evolution of molecular markers—just a matter of fashion?. *Nature reviews genetics*, 5(1), pp.63-69.
- Selkoe, K.A. and Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9(5), pp.615-629.
- Shafer, A.B., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W. and Wolf, J.B., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), pp.907-917.

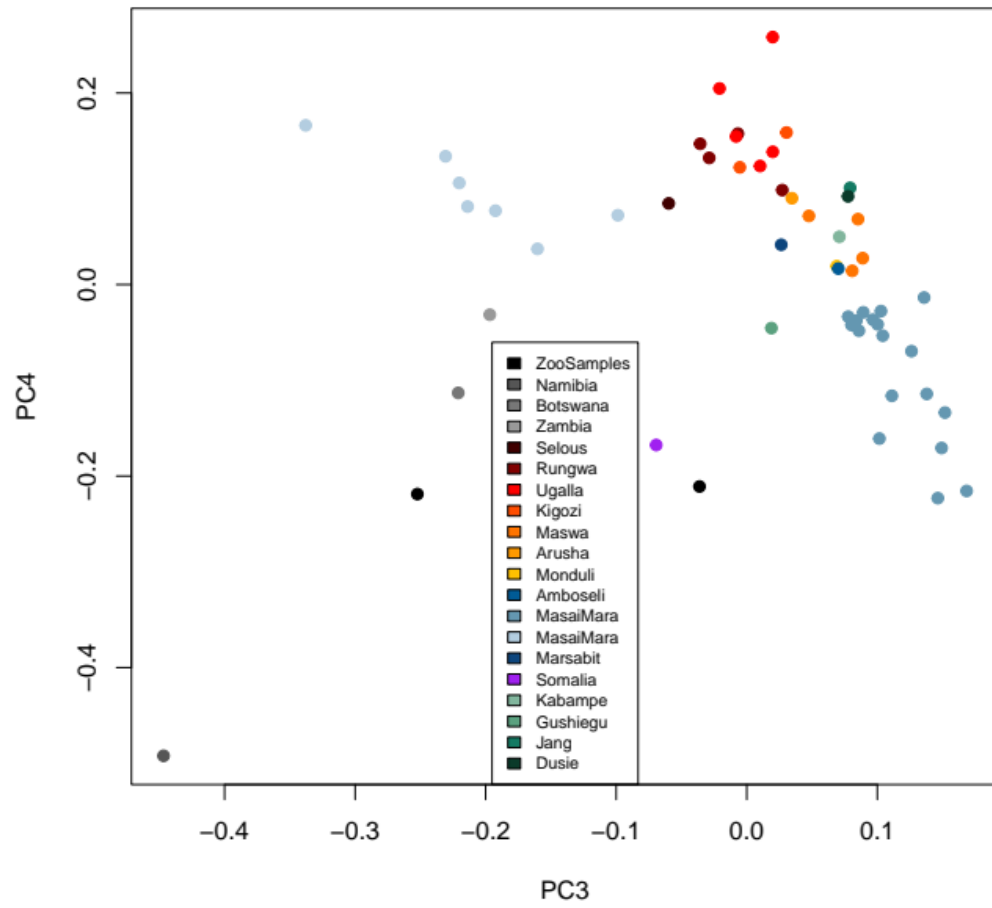
- Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J. and Willerslev, E., 2004. Rise and fall of the Beringian steppe bison. *Science*, 306(5701), pp.1561-1565.
- Shao, Y., Wang, X.B., Zhang, M.L., Liu, Y., Wang, S., Zhang, B.L., Yang, M.M., Yang, M.H., Jia, T., Pu, T.C. and Lu, Y., (2022). Long-read genome sequencing provides molecular insights into scavenging and societal complexity in spotted hyena *Crocuta crocuta*. *Molecular Biology and Evolution*, 39(3), p. msac011.
- Sheng, G.L., Soubrier, J., Liu, J.Y., Werdelin, L., Llamas, B., Thomson, V.A., Tuke, J., Wu, L.J., Hou, X.D., Chen, Q.J. and Lai, X.L., (2014). Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Molecular Ecology*, 23(3), pp.522-533.
- Smits, N., Jouvenet, O., Ambwene Ligate, F., Crosmay, W.G., Ikanda, D., Chardonnet, P., Fusari, A., Meganck, K., Gillet, F., Melletti, M. and Michaux, J.R., 2018. A genome-wide data assessment of the African lion (*Panthera leo*) population genetic structure and diversity in Tanzania. *PLoS One*, 13(11), p.e0205395.
- Szykman, M., Van Horn, R., Engh, A., Boydston, E. and Holekamp, K., 2007. Courtship and mating in free-living spotted hyenas. *Behaviour*, 144(7), pp.815-846.
- Trinkel, M., (2009). A keystone predator at risk? Density and distribution of the spotted hyena (*Crocuta crocuta*) in the Etosha National Park, Namibia. *Canadian Journal of Zoology*, 87(10), pp.941-947.
- Turner, A., Antón, M. and Werdelin, L., (2008). Taxonomy and evolutionary patterns in the fossil Hyaenidae of Europe. *Geobios*, 41(5), pp.677-687.
- Van Horn, R.C., Watts, H.E. and Holekamp, K.E., (2008). Do female hyaenas choose mates based on tenure?. *Nature*, 454(7201), pp.E1-E1.
- Varela, S., Lobo, J.M., Rodríguez, J. and Batra, P., (2010). Were the Late Pleistocene climatic changes responsible for the disappearance of the European spotted hyena populations? Hindcasting a species geographic distribution across time. *Quaternary Science Reviews*, 29(17-18), pp.2027-2035.
- Vellend, M. and Geber, M.A., 2005. Connections between species diversity and genetic diversity. *Ecology Letters*, 8(7), pp.767-781.
- Vissia, S., Wadhwa, R. and van Langevelde, F., (2021). Co-occurrence of high densities of brown hyena and spotted hyena in central Tuli, Botswana. *Journal of Zoology*, 314(2), pp.143-150.

- Watts, H.E., Scribner, K.T., Garcia, H.A. and Holekamp, K.E., 2011. Genetic diversity and structure in two spotted hyena populations reflects social organisation and male dispersal. *Journal of Zoology*, 285(4), pp.281-291.
- Werdelin, L., 1989. Constraint and adaptation in the bone-cracking canid *Osteoborus* (Mammalia: Canidae). *Paleobiology*, 15(4), pp.387-401.
- Werdelin, L. and Lewis, M.E., 2008. New species of *Crocota* from the early Pliocene of Kenya, with an overview of early Pliocene hyenas of eastern Africa. *Journal of Vertebrate Paleontology*, 28(4), pp.1162-1170.
- Werdelin, L. and Solounias, N., 1991. The Hyaenidae: taxonomy, systematics, and evolution. *In The Hyaenidae: taxonomy, systematics, and evolution* (pp. 1-104).
- Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L. and Hofreiter, M., 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(5), pp.1225-1237.
- Westbury, M.V., Hartmann, S., Barlow, A., Preick, M., Ridush, B., Nagel, D., Rathgeber, T., Ziegler, R., Baryshnikov, G., Sheng, G. and Ludwig, A., 2020. hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Science Advances*, 6(11), p.eaay0456.
- Woodruff, D.S., 2004. Populations, species, and conservation genetics. *Encyclopedia of biodiversity*, p.811.
- Yang, C., Li, F., Xiong, Z., Koepfli, K.P., Ryder, O., Perelman, P., Li, Q. and Zhang, G., 2020. A draft genome assembly of spotted hyena, *Crocota crocuta*. *Scientific data*, 7(1), p.126.
- Yirga, G., Ersino, W., De longh, H.H., Leirs, H., Gebrehiwot, K., Deckers, J. and Bauer, H., 2013. Spotted hyena (*Crocota crocuta*) coexisting at high density with people in Wukro district, northern Ethiopia. *Mammalian Biology*, 78(3), pp.193-197.

Supplementary Information

Appendix A:

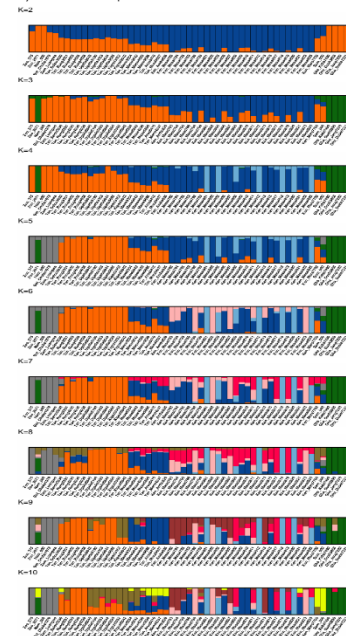




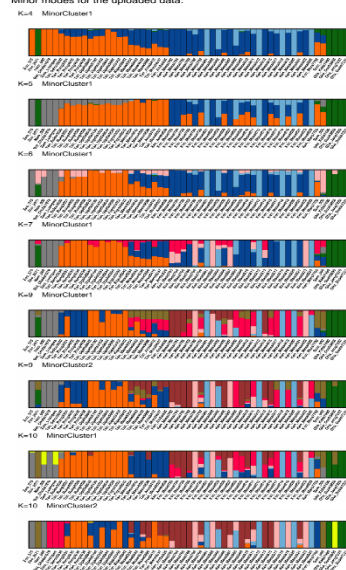
Appendix B:

CLUMPAK main pipeline - Job 1715253109 summary

Major modes for the uploaded data:



Minor modes for the uploaded data:



Division of runs by mode:

K=2 10/10
K=3 10/10
K=4 5/10, 5/10
K=5 8/10, 1/10
K=6 5/10, 5/10
K=7 5/10, 4/10
K=8 10/10
K=9 4/10, 3/10, 3/10
K=10 7/10, 2/10, 1/10

Appendix C:

