



UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE E BIOSCIENZE

CICLO XXXVII

The Tangled History of Baboons

Signals of hybridisation in the genome of *Papio* species

Coordinatrice:
Chiar.ma Prof.ssa Elena Maestri

Tutore:
Chiar.mo Prof. Cristian Capelli

Dottorando: Matteo Caldon

ANNI ACCADEMICI 2021/2022 - 2023/2024

Contents

List of Figures	V
List of Tables.....	VI
Abstract	IX

Introduction - Baboons, Hybridisations and the Species Problem

I - The Species Problem	1
I.I - <i>Species Tot Sunt Diversae Quot Diversas Formas ab Initio Creavit Infinitum Ens</i>	1
I.II - Towards the Modern Formulation of the Species Concept.....	2
I.III - Darwin and His “Dangerous Idea”	3
I.IV - The Modern Synthesis and the Biological Species Concept	5
I.V - Others Species Concepts	6
Morphological Species Concept (MSC)	6
Ecological Species Concept (ESC).....	7
Evolutionary Species Concept	7
Phylogenetic Species Concept (PSC)	7
I.VI – It Is Not About <i>How Many Angels Can Stand on the Head of a Pin</i>	7
II - Hybridisation.....	9
III - The Baboons: an Overview.....	11
III.I - Baboons in Biomedical Research	11
III.II - Baboons as a Model for Human Evolution.....	12
III.III - How Many Baboon Species	12
III.IV - How Baboons Became Baboons	15
III.V - Baboons Today	18
IV - Aim of the Work	21
V - References	21

Introduction - Overview of the Materials and Methods

VI - Dataset preparation.....	29
VII - Nuclear Genome Analysis	31
VII.I - Principal Component Analysis (PCA).....	31

VII.II - D-statistics or ABBA-BABA test	32
VII.III - f4-ratio test.....	33
VII.IV - ADMIXTURE	33
VII.V - Pairwise Sequentially Markovian Coalescent (PSMC)	34
VII.VI - hPSMC	34
VII.VII – Heterozygosity and Run of Homozygosity	34
VIII - References	35

Chapter 1 – Population Structure and Hybridisations in *P. ursinus*

1.1 - Introduction.....	37
1.2 - Material and Methods	41
1.2.1- Samples and Datasets	41
1.2.2 - Genome Mapping	43
1.2.3 - Variant Calling and Final Dataset	43
1.2.4 - Autosome Analysis	44
PCA	44
ADMIXTURE	44
Heterozygosity	44
D-statistics	44
f4-ratio	44
PSMC	44
1.3 - Results	45
1.3.1- Preliminary Analyses of Chacma Baboons at the Northern Border of <i>P. ursinus</i> Range: Zambia and Mozambique.....	45
1.3.1.1 - Chacma baboons in Gorongosa National Park, Mozambique	45
1.3.1.2 - Chacma baboons in Zambia	47
1.3.2 - Population Structure of Chacma Baboons	49
1.3.3 - Relationship Between <i>P. ursinus</i> and Neighbouring Species	53
1.3.4 - Demographic History	58
1.4 - Discussion.....	61
1.4.1 - <i>P. ursinus</i> Populations at the Northern Edge of Species Range: Zambia and Mozambique	61
1.4.2 - Hybridisations Events and Population Structure.....	62
1.5 - Conclusions.....	64

1.6 - References	64
-------------------------------	-----------

Chapter 2 - Hybridisation in Eastern Africa: *P. anubis* and *P. hamadryas*

2.1 - Introduction	69
2.1.1 - Science and Colonialism	69
2.1.2 - Museomics.....	71
2.1.3 - East African Baboons	72
2.2 - Material and Methods	75
2.2.1- DNA Collection, Extraction, Sequencing and Final Dataset	75
2.2.2 - Genome Mapping	77
2.2.3 - Variant Calling	77
2.2.4 - Autosome Analysis	77
PCA	77
ADMIXTURE	77
K-means clustering and D-statistics	77
F4-ratio	78
2.3 - Results	78
2.3.1 - Population Structure of Eastern African Baboons	78
2.3.2 - Hybridisation Events	83
2.4 - Discussion.....	84
2.5 - Conclusions.....	87
2.6 - References.....	87

Chapter 3 – *Theropithecus gelada* and the intergenic hybridisations with *Papio*

3.1 - Introduction	93
3.2 - Material & Methods	96
3.2.1- Sampling and Samples Dataset	96
3.2.2 - Genome Mapping	97
3.2.3 - Variant Calling and Final Dataset	97
3.2.4 - Autosome Analysis	98
Heterozygosity, ROH and Kinship	98
PCA	98
D-statistics	98

Admixture graphs	99
PSMC	99
AMY genes (Amylase)	100
Variation in Underground Food Consumption and Rainfall	101
3.3 - Results	102
3.3.1- <i>T. gelada</i> Population Structure and Genomic Variation	102
3.3.2- Demographic History of Gelada Populations.....	103
3.3.3- <i>Theropithecus-Papio</i> Hybridisations.....	104
3.3.4- Copy Number Variants.....	106
3.3.5- Variation in Underground Food Consumption and Rainfall	107
3.4 - Discussion.....	108
3.4.1- Population Dynamics and Climate Changes	108
3.4.2- Hybridisations	110
3.4.3- Climate, Diet and Local Adaptations	111
3.4.4- Implications for the Taxonomy and Conservation of <i>T. gelada</i>	112
3.5 - Conclusions.....	113
3.6 - References.....	114

Conclusions

Reticulate Evolution in Baboons	125
Web of Lineages: Gene Flow and Hybridisation in Baboons.....	125
Considerations on Conservation Status	126
Ethical Considerations Regarding Museum Collections	127
Future Directions.....	129
References	130
 Supplementary Table 1	 133
Supplementary Table 2.....	137
Supplementary Table 3.....	141
Supplementary Table 4.....	145
Appendix1	147

List of Figures

FIGURE 0.1 - Distribution of <i>Papio</i> species	15
FIGURE 0.2 - Phylogeny of Papionini tribe	16
FIGURE 0.3 - Main traits of baboon species	18
FIGURE 0.4 - Dated mitochondrial DNA phylogeny of the genus <i>Papio</i>	20
FIGURE 0.5 - D-statistics	32
FIGURE 0.6 - f4-ratio estimation	33
FIGURE 1.1 - Chacma baboon from Kruger National Park, South Africa	37
FIGURE 1.2 - <i>Papio ursinus</i> subspecies	38
FIGURE 1.3 - Map with the provenience of sampled individuals	41
FIGURE 1.4 - <i>Papio ursinus</i> skull (male, MSNG32161)	42
FIGURE 1.5 - Chacma baboons in Gorongosa NP	46
FIGURE 1.6 - Population structure of baboons in Zambia	47
FIGURE 1.7 - Proportion of chacma component in Zambian baboons.....	48
FIGURE 1.8 - Population structure of chacma baboons	50
FIGURE 1.9 - D-statistics comparing chacma baboons	51
FIGURE 1.10 - D-statistics comparing chacma and yellow baboon populations.....	52
FIGURE 1.11 - D-statistics comparing chacma and kinda baboon populations.....	53
FIGURE 1.12 - D-statistics comparing yellow and kinda with chacma baboon popula- tions.....	54
FIGURE 1.13 - Proportion of chacma component in chacma baboon populations.....	56
FIGURE 1.14 - f4-ratio tests for chacma and yellow baboon populations	57
FIGURE 1.15 - f4-ratio tests for chacma and kinda baboon populations	57
FIGURE 1.16 - Demographic history of Southern <i>Papio</i> populations	58
FIGURE 1.17 - Last gene flow timing between chacma baboon populations	59
FIGURE 1.18 - Last gene flow timing in Southern baboons.....	60
Figure 2.1 - Geographical map of Italian explorations in Eastern Africa (1870-1883) ...	70
Figure 2.2 - Eastern African baboons.....	72
Figure 2.3 -<i>Papio</i> skull	75
Figure 2.4 - Map with the locations of origin of modern populations and museum speci- mens of baboons	76

Figure 2.5 - PCA analyses of Eastern African specimens.....	79
Figure 2.6 - ADMIXTURE analyses on Eastern African baboons.....	81
Figure 2.7 - K-means clustering and D-statistics.....	82
Figure 2.8 - Proportion of olive component in olive baboon clusters.....	84
Figure 2.9 - Relationship between Eastern African baboons.....	85
Figure 3.1 – <i>Theropithecus gelada</i> (male).....	93
Figure 3.2 - Geographical origin of <i>T. gelada</i> samples.....	94
Figure 3.3 - Population structure of <i>Theropithecus gelada</i>	102
Figure 3.4 - Demographic history of gelada populations.....	103
Figure 3.5 - Gelada-baboons hybridisation events.....	105
Figure 3.6 - Allele sharing between Southern geladas and olive baboons	106
Figure 3.6 - CNVs and amylase gene copy number	107

List of Tables

Table 1 - Baboon (genus <i>Papio</i>) species classification.....	13
--	----

Abstract

Hybridisation represents a natural process whereby members of two distinct lineages come into contact and breed; when these events result in the production of fertile offspring, hybridisation can play a significant role in the evolutionary history of species. The majority of primate genera, including *Homo*, experienced similar events throughout their evolutionary histories. Among primates, baboons (genus *Papio*) represent a particularly notable case of interspecies (and possibly intergenera) hybridisation, with hybrid individuals observed in various regions across the African continent where the ranges of different baboon species overlap. Southern and Eastern Africa are two regions where baboons have been exposed to these phenomena, yet to be comprehensively investigated from a genomic perspective.

In order to investigate these events, new whole-genome sequences were generated from chacma, olive and hamadryas baboons, including museum specimens, commonly present in Southern and Eastern Africa, respectively. This thesis examines these novel genomes together with data already available from the literature.

In Southern Africa, hybridisation was detected between chacma baboons and their neighbouring species, the yellow and kindia baboons. The occurrence of population structure across chacma baboons and the extent to which hybridisation has shaped this structure were evaluated. Signals of hybridisation were also detected in Eastern Africa between olive and hamadryas baboons. The extent these interactions and others previously reported involving yellow baboons were reflected in the population structure of olive baboons was investigated. Finally, the evolutionary history of gelada (*Theropithecus gelada*), a primate endemic to Ethiopia that has also been observed to hybridise with olive baboons, was reconstructed. The population structure of geladas, their local adaptations, and the occurrence of genomic signals of hybridisation with baboon species were examined. The results presented in this thesis suggest that hybridisation is extensive in *Papio* species and that events of gene exchange have significantly affected local variation.

Introduction - Baboons, Hybridisations and the Species Problem

In this chapter, the issue of defining what constitutes a species in the biological realm will be explored, starting with a historical overview of how this problem has been addressed in the past. This introductory chapter will explore various concepts of species and the challenges they introduce. This *excursus* is fundamental to highlight why the definition of species is not just a theoretical issue: it is a real challenge that affects studies on Natural Science, the communication and understanding of their results and their implications on conservation politics. In this introduction, the concept of hybridisation will be also explored, as closely related to that of species. Besides being a subject of study, this concept is also one generating confusion in the definition of species. Finally, the group of animals (baboons) under investigation will be introduced presenting their variation and evolutionary history, highlighting how the problem of species concept applies to baboons and the significance of baboons as model for research.

I - The Species Problem

The definition of "species" is a fundamental concept in biology, with particular relevance for fields of study such as ecology, conservation, and taxonomy. Having a clear understanding of what is meant by a species is crucial for several reasons. Firstly, a precise definition allows for effective and consistent communication, ensuring that studies and data collected are comparable across time and useful on a global scale. Secondly, correctly identifying species is essential for the preservation of biodiversity. Misclassification can in fact lead to ineffective conservation strategies, either neglecting at-risk species or allocating resources to units that do not require protection. Finally, understanding the concept of species helps delineate evolutionary and ecological relationships among organisms, contributing to our knowledge of the life on Earth and the necessary measures to preserve it. But what do we mean by species?

I.I - *Species Tot Sunt Diversae Quot Diversas Formas ab Initio Creavit Infinitum Ens*¹

The term "species" derives from the Latin *species* (outward appearance), which corresponds to the Greek εἶδος, *eidos*. Plato used this term to refer to ideas or ideal forms. In Greek philosophy, it was related to the concept of genus (γένος, *genos*): compared to the latter, species represents a subdivision within a broader category. These are thus categories, a class of things, but also more simply a way to refer to the form of things themselves (Hey 2001). Referring to Platonic ideas, species were considered fixed and immutable forms, endowed with their own substance that distinguished them from other forms, from other species.

¹ From Latin: *there are as many diverse species as there are different forms created from the beginning by the Infinite Being*. Aphorism attributed to Carl Linnaeus

In the natural context, species were the forms of plants and animals, perfect and unchanging over time, preserved through the centuries as God (with the assimilation of Greek philosophy into Christianity) had created them at the beginning of the world. Each species was also incorporated into *scala naturae* or the great chain of being: a concept derived from Aristotelian and Neoplatonic ideas that influenced medieval and Renaissance thought. This vision holds that the world contains everything that could exist and that all qualities must be continuous, without sudden jumps (*natura non facit saltum*). It thus suggests that every living being is part of an ascending scale of perfection (Wilkins 2003).

Thus, fixism and essentialism were the two views surrounding the concept of species in the pre-scientific era. Fixism maintained that, following an initial creation, there were no other origins, and no creature had changed or gone extinct over time; essentialism, on the other hand, argued that living species are "natural types" and, as such, do not change.

I.II - Towards the Modern Formulation of the Species Concept

The first to formulate a concept of species applicable exclusively to reproductive forms (i.e., living beings) was the English naturalist John Ray (1627-1705). According to him, the criterion for determining species (of plants) is to observe the distinctive characteristics that are transmitted in seed propagation. He argued that variations arising from the same seed are accidental and not sufficient to distinguish a species. This principle could also be applied to animals: different species maintain their distinctive characteristics permanently and do not generate from one another. In his conception of species, the idea that there is an essence and a *scala naturae* remains (Wilkins 2003).

The Swedish botanist Carl Linnaeus (1707-1770) was the first to codify a system of nomenclature based on five ranks (classified in Regnum, Classis, Order, Genus, Species). The Linnaean ranks thus started from the species, which were named with the genus name followed by the species name. This system, with due integrations, is still in use today. Genus and species for Linnaeus continued to be fixed and real categories. In particular, he was convinced that genera were immutable entities created by God without the possibility of change; regarding species, however, towards the end of his life, he became convinced, through observation of the natural world, that some species were born through hybridisation and not created by God at the origin of the world (Hey 2001; Wilkins 2003).

During the same period, the French naturalist Georges-Louis Leclerc, Count de Buffon (1707-1788), made a significant contribution to the concept of species. He was a strong opponent of the Linnaean system, as he believed that variation among species was continuous and that changes in species were possible. He thought there were types of organisms approximately equivalent to Linnaeus's genera but that various species could degrade the original form, deviating from the true form, the true type. According to him, two animals belonged to the same species if, through copulation, they could maintain and perpetuate the similarities of the species; they belonged to different species if they could not produce offspring through copulation. Thus, for a species to be constituted, continuous, perpetual, and invariable reproduction was necessary. This idea of species was very similar to Ray's, but Buffon is credited with being the first to make reproductive isolation the test for determining whether two organisms should be

included in the same species. However, he also argued that only the organisms themselves were real and that species were merely conventions or convenient names for scholars (Wilkins 2003).

Jean-Baptiste de Lamarck (1744-1829), a disciple of Buffon, was among the first to combat species fixism, as he was the first to formulate an evolutionary theory of species. Lamarck's concept of evolution essentially introduced a temporal aspect to the *scala naturae*. He viewed each species as a distinct evolutionary lineage, originating through spontaneous generation from non-living matter and gradually ascending the great chain of being. Lamarck proposed the existence of a *feu étheré*, a physical energy that maintained organisms in their form and drove spontaneous generation from inorganic matter. He extended the concept of species to the mineral kingdom as well as the biological realm; however, he noted a key difference - mineral species lacked individuality, unlike plants and animals, which were individual entities capable of reproduction. Despite this, Lamarck, reflecting Buffon's influence, maintained that all classifications are arbitrary constructs of human thought, with only individuals truly existing in nature (Wilkins 2003).

I.III - Darwin and His “Dangerous Idea”

Charles Darwin (1809-1882) formulated the theory of evolution that is universally recognised today. In 1859, "On the Origin of Species" was published for the first time, marking a point of no return for the conception of life on Earth. The Darwinian theory of evolution was immediately a disruptive, uncomfortable, and very difficult idea to accept. For this reason, the philosopher Daniel C. Dennett spoke of "Darwin's Dangerous Idea" (the title of his 1995 book). Darwin undermined the logical explanation of the existence of God that dates back to the scholastic philosophy of Thomas Aquinas: complexity implies a designer; the universe is certainly highly complex; therefore, deductively, the universe has a designer (God or some other superior entity). With his theory, Darwin falsified this premise by showing that it is possible to have complex systems without resorting to a superior designer. Without an intelligent design that imposes order on nature, humans are no longer a special being made in the image and likeness of God, but just one of the many paths through which life has evolved. This revolutionary theory is based on two cornerstones: natural selection and common ancestry.

Natural selection is a mechanism of evolutionary change (not the only one) based on differential reproductive success among distinct types of biological entities. Natural selection itself is, in fact, this differential in reproductive success among distinct types of biological entities (Futuyma and Kirkpatrick 2017); this differential can have various natures and causes. It can be at the level of survival, or fertility, caused by a predator, a parasite, or temperature; but in any case, what happens is that some individuals perform better than others in terms of reproduction, and those are the ones that will pass their traits to future generations.

Common ancestry, on the other hand, is the concept that species multiply over time from a common ancestor. Tracing existing species back to a single ancestor does not imply that life had a single beginning from the inorganic world; there could have been other original forms of life, which, however, did not reach our time horizon. Additionally, tracing all current species back to a single ancestor does not mean that all other independent start-ups have no representation today: such an ancestor could have formed through different processes, for example

through some form of fusion/aggregation between the different start-ups, which therefore have current descendants.

These are two cornerstones of the theory, logically distinct and independent. In fact, we can conceive of a change in species through natural selection that does not produce a multiplication of species; at the same time, we can imagine a process of species multiplication not produced by natural selection. Thus, saying they are logically distinct means that we can conceive of evolutionary change mechanisms in which we observe one of these elements and not the other. However, from another point of view, they are evidentially connected: natural selection and common ancestry are closely connected at the level of evidence, the way in which each of the two concepts is supported or backed by observations and arguments. Natural selection can cause divergence, and this happens when the selective regimes of one population differ from those of another population. Adaptation to different conditions leads to divergence. Thus, natural selection explains why current forms trace back to one or a few original progenitors; however, it does not imply it (Sober and Orzack 2003).

With the formulation of his theory, Darwin definitively surpassed any fixed and essentialist idea related to the concept of species. Although he never provided a formal definition of species, Darwin revolutionised the traditional view of species, considering them as descendants "with modifications" from common ancestors and comparing them to "branches in lines of descent" (de Queiroz 2011). He argued that species, which today appear discrete, were not so in the past as they diverged from a common ancestor. Current species are state relationships in the flow of morphological change, which will continue to accumulate with geological time. In Darwin's view, the temporal discreteness between species was completely shaken, and even contemporary discreteness would seem to crumble if sampled near the point of divergence (Eldredge 1993). The idea that species are segments of lines of descent is fundamental to the modern concept of species and has influenced most subsequent views (de Queiroz 2011).

Despite his innovative evolutionary conceptualisation, Darwin's concept of species retained an element of the previous taxonomic tradition: the idea of species as a rank within the taxonomic hierarchy. Darwin applied a conventional criterion for defining species, focusing on the degree of difference and the absence of intermediates between groups. These propositions were fundamental to arguing that varieties are incipient species and that taxa at all levels of the taxonomic hierarchy can be explained by the process of descent with modification (de Queiroz 2011).

Darwin also extended the idea that varieties and subspecies represent intermediate stages in the evolution of species, concluding that varieties are species in formation. If varieties are indeed incipient species, then they represent species in the early stages of their development. This implies that the segment of a line of descent conceived as a species extends much further back towards its origin (Eldredge 1993).

Darwin viewed variation within species as potential for future evolutionary change, though with an uncertain outcome. He believed that species confer "permanence" to morphological changes through reproductive discontinuity, which allows these changes to be incorporated into the broader phylogenetic flow. Although Darwin considered different perspectives on the issue of species discreteness, he ultimately replaced the notion of static and discrete entities with a concept of continuous biotic flow (Eldredge 1993).

After Darwin, many post-Darwinian biologists saw species as minimally differentiated collections of organisms. However, in the late XIX and early XX centuries, some biologists highlighted the significance of reproductive isolation in closely related species. The ongoing debate about the distinctness of species in space and time has always been connected to the issue of tokogeny (the interconnection of genetic lines through sexual reproduction) and its impact on morphological similarities and differences (Eldredge 1993).

Throughout these debates, species have been considered in various ways: as entirely reproductive entities, morphologically defined subsets within reproductive entities, non-real entities, discrete spatially, temporally, both or neither. Despite the apparent chaos, there is a logical structure and order to these various positions in the debates on species (Eldredge 1993).

I.IV - The Modern Synthesis and the Biological Species Concept

Following a phase of eclipse of Darwinism at the beginning of the XX century, in the 1930s and 1940s, Darwin's ideas returned to prominence in the scientific community, later being re-interpreted in a modern key. This resurgence was particularly due to the rediscovery of Mendelian genetics and its integration into Darwinian evolutionary theory, marking the beginning of the Modern Synthesis of the Theory of Evolution (named after Julian Huxley's 1942 book "Evolution: The Modern Synthesis," one of the main characters of that period for the biological sciences). This new phase added additional elements to the central core of Darwinian theory such as population genetics and Mendelian inheritance. We can affirm that Modern Synthesis is a body of evolutionary models based on the measurement of allele and genotype frequencies; it is possible to trace changes in the genetic makeup of organisms to understand evolutionary changes at the phenotypic level.

These allele-based and genotype-based models are built on certain implicit assumptions: organisms have genomes; genomes are sets of genes, each of which can be present in alternative versions called alleles; the transmission of these genes occurs through Mendelian inheritance; the genotype determines the phenotype. In this way, the theory shifts from focusing on phenotypes to genotypes.

In this context, the biological species concept (BSC), the most well-known and generally used concept to define what is or is not a species, was born. This concept was developed to integrate reproductive discontinuity with Darwin's evolutionary view, which emphasised continuity among species. According to Dobzhansky (1937), Darwinian theory focuses on the origin, maintenance, and modification of organism traits, viewing species as groups of similar organisms, with the action of natural selection producing a continuum, an uninterrupted spectrum of morphological variation. However, according to Dobzhansky, in nature, species exist as more or less discrete entities; therefore, he introduced a second theme into evolutionary biology alongside the diversity of forms: discontinuity (Eldredge 1993).

Dobzhansky (followed by Mayr) believed that discontinuity was a primary result of evolution and that it allowed species to concentrate on adaptive peaks. He also viewed speciation through the action and evolution of isolation mechanisms as a process that drives two incipient species to become fully separate, better focusing on the adaptation of organisms to their respective

niches or adaptive peaks. This perspective is rooted in the idea that natural selection should produce a continuum of adaptive diversity, which led Dobzhansky and Mayr to adopt a reproductive concept of species (Eldredge 1993).

Dobzhansky and Mayr definition of species emphasized reproductive relationships within species and the lack of such relationships beyond the boundaries of a species. The BSC views species as reproductive communities whose members are more or less similar to each other, but much more so than they are with members of other reproductive communities. Reproductive discontinuity is thus the first necessary step to have a phenotypic discrepancy between different groups (Eldredge 1993).

Hence, the definitions given by Dobzhansky and Mayr (Dobzhansky 1937; Mayr 1942) were almost identical and can be summarised as follows: "Species are groups of actually or potentially interbreeding natural populations, reproductively isolated from other such groups". In other terms, a species is a group of natural populations that interbreed with each other and live in a particular environment but are incapable of crossing with other groups (Aldhebiani 2018).

The BSC can encompass two other species concepts: the isolation concept and the recognition concept. Both are based on reproductive barriers observed between different forms. Specifically, the isolation concept emphasizes that reproductive isolation is the key mechanism for the discontinuity that can be observed between species, whereby each species is reproductively isolated from others, preventing the mixing of genes and traits with other species. The recognition concept, on the other hand, stresses reproductive cohesion as the factor responsible for species continuity (Aldhebiani 2018).

Despite its simplicity and obviousness in most cases, the BSC presents a significant number of situations where this concept is not applicable. Firstly, it is not usable to recognise different species of asexual organisms, such as bacteria. Additionally, it is impractical in cases of allopatric populations, thus geographically isolated, or in the presence of continued hybridisation and introgression events over time (Aldhebiani 2018).

I.V - Other Species Concepts

In modern science, the concept of species has been defined in many different ways beyond the BSC by taxonomists over the past sixty years. Some of these concepts are phenetic, while others are phylogenetic. Phenetic concepts apply a classification based on the similarity of current characteristics among organisms. Phylogenetic concepts, on the other hand, are based on cladistics, thus classifying groups of individuals based on how the characteristics of the organisms have appeared in their current state (Aldhebiani 2018). Below are some of the most relevant modern concepts of species.

Morphological Species Concept (MSC)

It defines species as "a community, or a number of related communities, whose distinctive morphological characters are, in the opinion of a competent systematist, sufficiently definite to entitle it, or them, to a specific name" (Regan 1926). It can be applied to both asexual and sexual organisms and can be used to the fossil record. Unfortunately, in some cases the morphological features are not objective and depend on the judge of "experts" for key traits. The fundamental

problem with this species concept is that some species are morphologically identical but belong to different evolutionary lineages (Aldhebiani 2018).

Ecological Species Concept (ESC)

Van Valen (1976) defined a species as "a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range." This concept, therefore, is based on the idea that when two organisms are similar to each other, they tend to overlap ecologically and compete, and are thus likely to belong to the same species. The issue with this concept, however, is that it is not always possible to determine the degree of ecological competition between entities, nor is it easy to determine when the process of dividing divergent forms into new species stops (Aldhebiani 2018).

Evolutionary Species Concept

This concept characterises a species as a "single lineage of ancestor-descendant populations of organisms which maintains its identity from other such lineages [in space and time] and which has its own evolutionary tendencies and historical fate" (Wiley 1981). It was developed to accommodate asexual organisms and extinct species to which the BSC could not be applied. However, a challenge with the evolutionary species concept emerges when gaps in the fossil record create artificial boundaries between species, particularly for those undergoing gradual evolution in size or shape over time (Aldhebiani 2018).

Phylogenetic Species Concept (PSC)

Under this concept species are defined as individuals who exhibit a high degree of similarity in numerous distinct traits, forming monophyletic clusters based on shared phenotypes. In essence, species are seen as groups of organisms that descend from a common ancestor. This concept integrates character-based concepts, which emphasise the presence of shared traits, and history-based concepts, which emphasise the degree of relationship of an organism with other previously characterised organisms. Unlike the Biological Species Concept (BSC), this concept can be applied to asexual and sexual populations, as well as allopatric populations. However, it is not always possible to accurately reconstruct the evolutionary history of a group of individuals (Aldhebiani 2018).

I.VI - It Is Not About *How Many Angels Can Stand on the Head of a Pin*²

As has been noted, the conceptual debate on what is a species is very complex and lengthy. There is no true universal definition, and different concepts are used depending on the situation. This variability, however, can lead to confusion. Clarity is crucial because the conceptual foundation has a direct impact on practical science. For instance, the relevance of clear species definitions is significant in biology and related fields. It is important to recognise that the choice of species concept can influence the outcomes of scientific studies and conservation strategies. In biomedical research, precise species identification is crucial for understanding and making clear what is being worked on, whether the individuals tested are truly representative of the

² The phrase, now used as a metaphor for intellectually and practically pointless debate topics, originates from XVII century Protestants, who used to mock the futile debates of medieval scholastics such as Thomas Aquinas.

entire species under study or just a specific group (if a broad species concept is used). An issue of relevance arises in the species *Macaca fascicularis*, frequently utilised in biomedical research, which has a large range in South-West Asia. The populations from different geographic areas are characterized by different genetic compositions, so researchers must verify the subjects' origin, ancestry, and lineage to minimise risks in biomedical experiments arising from genetic variability due to close kinship and mixed ancestry. These factors have the potential to obscure treatment effects (Kanthaswamy et al. 2013).

Similarly, in conservation, accurate species delineation can determine the allocation of resources and the prioritisation of efforts to protect biodiversity. A group may be at risk of extinction, but if considered part of a broader species that is not at risk, then conservation programs for that group will not be implemented. Conversely, if that group is considered a species, its conservation status will change, and it will receive more attention for its preservation. An example is the Tapanuli orangutan: in 2013, this population was considered a separate species (Nater et al. 2017), and it was realised that it is critically endangered (only 800 extant individuals), making it essential to focus more on the conservation of this population/new species. The use of museomic data, which involves genetic information extracted from museum specimens, also relies heavily on clear species concepts to draw meaningful conclusions about evolutionary history and species interactions. If an individual's nomenclature is not precise, there is a risk of using an individual, thinking it is something when it is something else, leading to problems in interpreting the data and rendering that sample less informative (Kitchener et al. 2020).

Therefore, the adoption of a clear and consistent species concept is not merely an academic exercise but a practical necessity. It ensures that research findings are robust, conservation strategies are effective, and historical data from museum collections are accurately interpreted. This highlights the interconnectedness of theoretical debates and practical applications, emphasizing the need for careful consideration and transparent communication in the scientific community.

Although there is still no single agreed-upon species concept, we can summarize some operationally useful general concepts in evolutionary biology. Firstly, species are not arbitrary but objectively distinct groups, whose diversity is independent of human perception. Biodiversity is naturally clustered into distinct and recognisable groups. However, this clustering is not always well-defined due to the dynamic nature of species and their lack of an essence. On the contrary, species evolve, but this process is gradual, implying that significant differences arise from the accumulation of small differences within the population. Speciation can occur relatively quickly or very slowly, but there will always be a period of uncertainty where evolutionary separation is not yet complete (Fitzpatrick et al. 2015).

Reproductive isolation, as previously discussed, is linked to maintaining a certain divergence between two groups, depending on the tension between processes that lead to homogeneity like gene flow, and those that promote divergence such as mutations and genetic drift. There is no doubt that reproductive isolation is essential for maintaining diversity among populations. Different forms that hybridise freely, therefore, will not remain distinct in subsequent generations. There will thus be non-homogeneous and stable population structures, previously distinct groups forming hybrid zones, or incipient groups still connected by gene flow (Fitzpatrick et al. 2015).

II - Hybridisation

“Hybridisation” is a process in which individuals from genetically distinct and divergent lines interbreed and produce fertile offspring, playing a creative role in evolution (Arnold and Meyer 2006; Cortés-Ortiz 2017; Taylor and Larson 2019). These events are relatively rare; however, it is estimated that about 10% of primates are (or have been) involved in such processes (Arnold and Meyer 2006; Zinner, Arnold, and Roos 2011; Cortés-Ortiz 2017).

Hybridisations can be ongoing (resulting in intermediate morphotypes between two species with a morphocline in the hybridisation area) or ancient. In the latter case, their detection is complex, requiring appropriate genomic tests for the taxa involved (Zinner, Arnold, and Roos 2011). One of the most famous cases of ancient hybridisation involves our species. Genomic studies following the ancient DNA revolution (which awarded the 2022 Nobel Prize in Medicine to one of its leading figures, Svante Pääbo) have shown that there was introgression of at least two extinct archaic human forms (Neanderthals and Denisovans) into non-African populations of *Homo sapiens*. Today, about 2% of the genome in non-African populations comes from Neanderthals, and up to 6% in Oceanian populations from Denisovans (Reich 2018).

In nature, hybridisations occur in geographically restricted areas known as hybrid zones (Cortés-Ortiz 2017). These zones often exhibit a cline in variation maintained by the balance between dispersion and selection against hybrids. In these zones, the population transitions from pure form A to pure form B through the hybrid form. In other cases, the hybrid zone is patchy, meaning the transition between form A and form B is not clinal (Harrison and Larson 2016).

Depending on the process that leads to the formation of the hybrid zone, two categories are distinguished: primary hybrid zones develop *in situ* when natural selection and environmental differences alter allele frequencies in contiguous populations. Secondary hybrid zones form when two genetically differentiated, previously allopatric populations expand their ranges to meet and subsequently interbreed (secondary contact). Most hybrid zones involving primates seem to fall into the latter category (Cortés-Ortiz 2017).

In a classic scenario of allopatric speciation, a population splits into two populations that develop their independent evolutionary histories in distinct geographical areas, leading to speciation. If secondary contact occurs after a reasonably long period, genomic, physiological, or behavioural differences prevent reproduction (supporting the BSC). However, if insufficient time has passed for reproductive barriers to form, there can be gene flow between the two species, producing first-generation (F1) hybrids (Zinner, Arnold, and Roos 2011). These hybrids have an equal proportion of the genome from both parental lines, the mitochondrial genome from the maternal line and the Y chromosome from the paternal line (Cortés-Ortiz 2017). In many cases, hybrids are sterile or have significantly decreased fitness when compared to the parental generation (again fitting the BSC case). Nonetheless, F1 hybrids can be fertile and viable, leading to backcrossing phenomena where hybrids mate with one or both parental populations (Zinner, Arnold, and Roos 2011). After many generations of repeated hybridisation events and backcrossing, individuals with varying levels of genetic admixture are generated. Those events iterated with one or both parental forms can produce a wide range of innovative genotypes among individuals in the hybrid zone. Natural selection then eliminates less fit

genomic combinations while favouring more performant ones (Zinner, Arnold, and Roos 2011). As a result, hybridisation can dissolve species boundaries, blending two gene pools, weakening reproductive isolation, and enhancing diversity in one or both parental lineages through introgression (when alleles transfer from one gene pool to another via admixture; Cortés-Ortiz 2017).

Hybrid zones act as bridges through which divergent genomes meet and swap parts of their material. Allopolyploidy is common in plants, where parental genome duplication leads to immediate incompatibility with parental species, establishing a unique niche. In contrast, homoploidy is more common in animals, where hybrid progeny have the same chromosome number as the parental generation (Abbott, Barton, and Good 2016; Taylor and Larson 2019). The amount of genetic introgression in hybrid forms can depend on the divergence level of the parental forms before hybridisation and the performance of exchanged genomic regions relative to the receiving genomic background (Cortés-Ortiz 2017). Neutral genes can be exchanged without affecting hybrid fitness, remaining phenotypically silent in the host genome, while other genes may have negative effects and be eliminated by selection in subsequent generations, showing few or no introgression signals. Conversely, introgressed genes can positively affect hybrid fitness, providing an adaptive advantage that allows them to quickly spread within the hybrid population. The mixing of two originally divergent genomes within hybrid zones through events of hybridisation, recombination and backcrossing creates opportunities for selection to influence new gene combinations. This process enables certain genomic regions to merge while restricting the integration of others. Consequently, the genomes of hybridising species exhibit semi-permeability to introgression, with the extent of this permeability varying according to the specific properties of different genomic regions (Cortés-Ortiz 2017).

Hybridisations pose a significant challenge in recognising populations as distinct entities, which has crucial implications for their conservation status (Zinner, Arnold, and Roos 2011). When gene flow between species extends beyond the F1 hybrid generation and continues in subsequent generations, a complex issue arises. In conservation, hybridisation has almost always been viewed negatively, especially concerning populations of small dimensions belonging to threatened or endangered species. It has been hypothesised that these small populations could be lost if extensively hybridised with more abundant congeners. Additionally, human activities that lead to habitat fragmentation or population reduction can promote or intensify hybridisation. For instance, in a small, fragmented habitat where two closely related species coexist and only a few males of one species survive, these males may only have the option to mate with females of the other species. Finally, hybridisation can significantly impact conservation strategies for certain taxa. Even though unique and ecologically important hybrid forms may emerge, they are often not recognised as distinct taxonomic entities and, consequently, are unlikely to receive legal protection (Zinner, Arnold, and Roos 2011).

The question, therefore, is whether hybrid populations should receive the same attention as pure populations. Even if considered distinct entities, such populations might have the characteristics to be protected and subject to conservation actions. For this reason, biodiversity protection policies should be reviewed to be more inclusive and closer to evolutionary units below the species level (Fitzpatrick et al. 2015).

III - Baboons: an overview

Baboons in Ancient Egypt (as early as the IV millennium Before the Common Era) were depicted as the personification of the underworld God Babi. They were sometimes also portrayed as the reincarnation of the lunar and wisdom God Thoth. Numerous depictions show a male baboon in a seated position, hands on knees, with a lunar disk above its head. In various sites, mummified remains of baboons have been found. Since baboons are not native to the region, they must have been imported from their original habitats and kept in captivity. Many of these remains showed limb deformities due to a lack of sunlight and poor nutrition, indicating the precarious conditions under which they were kept despite their sacred status (Dominy et al. 2020; Van Neer et al. 2023).

III.I - Baboons in Biomedical Research

The use of baboons (genus *Papio*) as research animals began in April 1956 when Dr. Nicholas T. Werthessen, a researcher at the Southwest Foundation for Biomedical Research (SFBR; now the Texas Biomedical Research Institute in San Antonio, USA), was visiting Dr. Russell L. Holman, at the Louisiana State University School of Medicine. Holman and Werthessen were discussing issues related to heart and arterial diseases when an assistant offered Holman an aorta nearly 60 cm long, filled with diseased areas. The aorta came from a 16-year-old baboon from the Audubon Park Zoo in New Orleans. The significance of this discovery was that previously no animal was known to develop arterial diseases similar to those in humans. This serendipitous discovery of an atherosclerosis-like lesion in a 16-year-old baboon led the researchers to consider the baboon as a potential model for investigating atherosclerosis (VandeBerg, Williams-Blangero, and Tardif 2009). Thus began the interest in biomedical research for these animals, and in 1960 the first group of baboons was sent from Kenya to the Southwest Foundation in Texas, where the first colony of baboons was established (Fischer et al. 2019). Today, that colony is managed by the Southwest National Primate Research Center (SNPRC) and includes more than 1,000 baboons, the largest in the world. Since then, the use of captive baboons as a model for biomedical research has been second only to that of macaques (that have been used for much longer in this field) and has been supported by extensive field studies on wild baboon populations in Africa. Understanding the connections between fitness and health in baboons in their natal environments, including their natural levels of genotypic and phenotypic variation, is essential for contextualizing the results of captive studies. Furthermore, studying the evolution and life history of these animals in their natural habitats enables the relevance and accuracy of results that emerged from studies conducted on captive populations (Fischer et al. 2019).

The value of this genus is often highlighted by the substantial variation in life history traits, ecologies and social systems both within and among baboon species. This diversity makes baboons a very good model organism for studying a broad spectrum of fundamental biological processes, including behavioural and physiological adaptation, speciation driven by gene flow and hybridisation. Consequently, the baboon as a model serves as an important complement to other mammalian model organisms. (Fischer et al. 2019).

III.II - Baboons as a Model for Human Evolution

Introgression and hybridisations have become central dynamics in the research of human origins, especially concerning past events of gene flow between anatomically modern humans and archaic human populations. This issue is central to debates concerning the most recent human evolutionary history and the two models of modern human origins: the multiregional and the recent African origin (or replacement) model. The two models present contrasting views on the role of admixture. The multiregional model suggests that admixture was crucial to maintain modern humans as unique specie, given that (with this model) they evolved from regional hominins like *Homo erectus* (Wolpoff et al. 1984). On the other hand, the replacement model argues that modern humans emerged recently in Africa and spread across the globe, replacing local archaic species with little to no admixture (Stringer & Andrews, 1988).

Recent advancements in ancient DNA technology have shed new light on these theories. The sequencing of the Neanderthal genome revealed 1-2% admixture between these hominins and Eurasians, with no evidence of admixture between Africans and Neanderthals (Green et al. 2010). This discovery challenges both models: equal admixture levels between all Eurasians and Neanderthals were not predicted in the multiregional model, while the replacement model expects minimal or no admixture between modern humans and archaic hominins that evolved outside of Africa.

The discovery of another archaic human, the Denisovans, in Denisova Cave (Siberia) and their DNA analysis highlight the contribution of Denisovans to modern human populations in Oceania and South-East Asia (Reich et al. 2010, 2011). These discoveries suggest multiple episodes of gene flow with archaic human forms, influencing the genomes of now-day human populations.

To evaluate the human evolutionary dynamics that cannot be fully understood through the human model, nonhuman primates are particularly useful. Jolly (2001) has suggested that baboons could serve as excellent models for understanding human population structure before the "Out of Africa" migration of modern humans. Baboons are valuable analogues for humans because they share numerous homologous traits due to their phylogenetic proximity, yet they also exhibit many nonhomologous characteristics in their morphology, behaviour, ecology, and population structure. Notably unlike the great apes, which are phylogenetically closer to the genus *Homo*, baboons share the same savanna environments and face similar ecological pressures experienced by early African human forms.

III.III – How Many Baboons Species

Baboons are classified as morphologically distinct allotaxa (geographically-replacing forms) that are not reproductively isolated. Baboon allotaxa frequently encounters one another, where neighbouring groups occupy similar ecological niches but display phenotypic and behavioural differences. These differences may reflect adaptations to slightly varying habitats. Furthermore, these allotaxa often show distinct social organization patterns, some of which are genetically influenced.

The initial distinction among the various forms of baboons was made by observing fur characteristics. These traits are fairly stable, although they can vary slightly between populations, with

the notable exception of those at the edges of where different forms range showing clear signs of secondary hybridisation (Jolly 1993). All recognisable forms constitute together a single continuous series where neighbouring populations differ along a cline, showing continuous variation from one form to another (Jolly 1993). For this reason, this group of animals serves as an excellent example to illustrate the issue of what constitutes a species. There is a wide diversity of traits, and depending on the operational concept of species used, the number of emerging species varies greatly. There is also the issue regarding hierarchy among forms; is it correct to elevate some forms to species while relegating others to subspecies?

Table 1: Baboon (genus *Papio*) species classification. PSC (Phylogenetic Species Concept), ESC (Ecological Species Concept), BSC (Biological Species Concept).

Common name	PSC (6 species)	PSC (5 species)	ESC	BSC
hamadryas	<i>hamadryas</i>	<i>hamadryas</i>	<i>hamadryas</i>	<i>hamadryas</i>
Guinea	<i>papio</i>	<i>papio</i>	<i>cynocephalus</i>	
olive	<i>anubis</i>	<i>anubis</i>		
Heuglin's				
Ibean yellow	<i>cynocephalus</i>	<i>cynocephalus</i>		
typical yellow				
Kinda	<i>kindae</i>			
gray-footed	<i>ursinus</i>	<i>ursinus</i>		
Ruacana				
typical chacma				

Since interbreeding has been observed between pairs of parapatric forms (as discussed later), we should recognise a single species of baboon according to the Biological Species Concept (BSC). Jolly (1993) proposed that baboons represent a single polytypic species (*Papio hamadryas*), including 10 recognisable phylogenetic subspecies (Guinea, Anubis or Olive, Hamadryas, Heuglin's, Ibean yellow, typical Yellow, Kinda, Gray-Footed, Ruacana and typical Chacma; Table 1). His judgment is drawn from the observed discrepancies between "phenostucture" and "zygostructure" in baboons. Phenostucture refers to observable characteristics of organisms at any level, including social systems, phenotype and genotype. Zygostructure instead, focuses on the likelihood of forming viable zygotes and encompasses all aspects of population structure in genetic terms, such as rates of gene flow, cross-mating, and the extent of positive and negative assortative mating for phenotype or inbreeding avoidance. According to Jolly, the zygostructure of baboons consists of four isolated populations among which gene flow is restricted: the two isolated olive baboon populations in the Sahara Desert, the Arabian hamadryas population, and the rest of all other baboon populations in Africa (Jolly 1993; Zinner et al. 2011).

A second alternative to the classification problem of baboons included two species (Smuts et al. 1986; Richard 1985): *P. hamadryas* and *P. cynocephalus*. The first was represented only by the form resident in East Africa and the Arabian Peninsula characterised by a different social system, ecology and morphology compared to all other forms; while the latter encompassed all other existing forms on the African continent and was thus subdivided into numerous subspecies. This classification was primarily formulated based on the ecological concept of species, distinguishing between "desert baboons" (*P. hamadryas*) and "savanna baboons" (*P. cynocephalus*). The problem with this potential solution lay in the fact that the two forms whose

hybridisation was well-documented and successful (hamadryas and olive) in the Awash region of Ethiopia were divided into two distinct species (Phillips-Conroy & Jolly, 1986), while have been grouped as belonging to the same species forms whose contacts have never been observed and for geographical reasons could never occur (Guinea and kinda, or Guinea and chacma; Jolly 1993; Zinner et al. 2011). Moreover, the desert-savanna baboon species dichotomy is not entirely accurate, as baboons have been observed living in arid environments in the areas surrounding the Namib and Kalahari deserts in southern Africa (Hamilton 1986).

To address the issues arising from previous proposals, a third alternative was formulated aiming to elevate the status of the five main forms (hamadryas, olive, Guinea, yellow, and chacma) to species level (Groves 2001; Grubb et al. 2003). This approach aimed to resolve the conceptual problem of uniting what could not be united based on zygostructure, and separating what did not need to be united in the context of the entire baboon population. However, the flaw in this model is that it introduces boundaries that do not exist in nature; for example, *P. anubis heuglim* and *P. cynocephalus ibeanus* are extremely similar but are recognised as belonging to different species. This model, like the two-species model, cannot be supported by the Biological Species Concept (BSC); however, in this case, it was supported by the Phylogenetic Species Concept (PSC), as five phylogenetic groups were observed. With this solution, populations presumed to be the major forms or the typical forms are those found at the ends of the cline. Therefore, the end forms are the most extreme, but the populations connecting them are real and phenotypically stable. This scheme also implies a well-defined evolutionary history, suggesting that these species are the result of initial radiation and that subspecies formed subsequently through internal diversification and hybridisation (Jolly 1993; Zinner et al. 2011).

Further complicating the overall picture is the issue of *Theropithecus gelada* (this will be discussed further in the third chapter). Hybridisation events between this species and olive baboons in Ethiopia have been suggested, but these primates have been assigned not only to different species but also to different genera (Dunbar and Dunbar 1974). This example highlights the limitations of using the reproductive isolation criterion of the BSC to define species boundaries, as it can lead to classifications that do not align with traditional species taxa. However, the taxonomic categorisation is less significant than the observed mismatch between zygostructure and phenostructure. What stands out is that an olive baboon in central Ethiopia is more likely to successfully reproduce with a gelada than with another olive baboon from Western Africa, the isolated population in the Sahara, or a South African chacma baboon. This discordance between zygo and phenostructure could be quite ancient. The gelada belongs to a clade that, based on abundant fossil records, has been phenetically different from *Papio* forms for at least 4My, during which time it has undergone significant radiation (Jolly 1972; Jablonski 1993). Throughout most of this time, various species of *Theropithecus* were widespread across the African plains, particularly in the Southern regions far beyond the current *T. gelada* range in Ethiopian highlands, and in some cases sympatric with *Papio*. This scenario suggests the intriguing possibility that *Theropithecus* and *Papio* may have exchanged genes during their respective internal cladogenesis and adaptive radiations. This case powerfully illustrates the potential decoupling of zygostructure and phenostructure divergence by highlighting the inadequacy of the BSC in defining the species of this animal group. This case study underscores the need to integrate multiple taxonomic systems to account for population structures and their evolutionary history (Zinner et al. 2011).

Zinner and colleagues (2011) proposed, following suggestions by others and building on the five-species classification, to elevate a sixth form to species rank, *P. kindae*. The reasons provided relate to the fact that these baboon groups are effectively different from all others. In fact, they are smaller in size, show less pronounced sexual dimorphism, present cranial morphometry significantly different from *P. cynocephalus* (the species they were initially considered a subspecies of), and exhibit behavioural peculiarities. Furthermore, analysis of their mitochondrial genome reveals that they constitute a monophyletic group (Zinner et al. 2011).

Despite the aforementioned issues, for ease of communication, six species are currently recognised under the PSC: *P. ursinus* (chacma baboon) living in Southern Africa; *P. cynocephalus* (yellow baboon) inhabiting from Southern Somalia to Mozambique; *P. kindae* (Kinda baboon) found in Angola, Zambia, and the Southern part of the Democratic Republic of Congo; *P. anubis* (olive baboon) extending its range from Northern Democratic Republic of Congo through Kenya and Tanzania, passing through Sierra Leone in the West and Eritrea in the East; *P. papio* (Guinea baboon) found between Sierra Leone, Mauritania, and Senegal; and *P. hamadryas* (hamadryas baboon) inhabiting the entire Horn of Africa and the South-Western part of the Arabian Peninsula (Zinner et al. 2011; FIGURE 0.1).

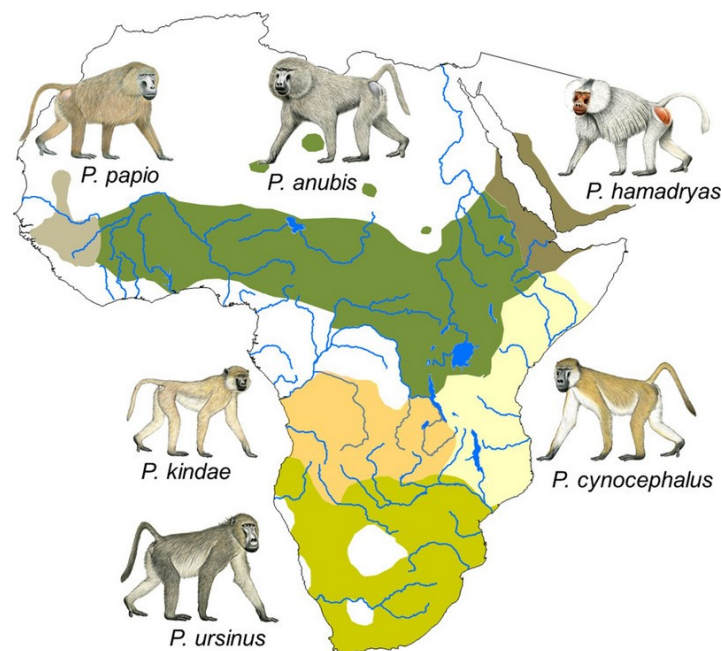


FIGURE 0.1 - Distribution of six *Papio* species. From Fisher et al. 2019. Male baboon drawings by Stephen Nash.

III.IV - How Baboons Became Baboons

All existing baboons belong to the genus *Papio*; this genus is part to the Papionini tribe and the Cercopithecidae family. The now-day extant genera belonging to this tribe are *Mandrillus*, *Cercocebus*, *Lophocebus*, *Rungwecebus*, *Theropithecus* and *Papio*. The primary subdivision within the tribe distinguishes two clades: *Mandrillus* and *Cercocebus* and all other genera. The division of the second clade is a subject of significant debate (Zinner, Arnold and Roos 2009; Liedigk 2014; Pugh and Gilbert 2018). Current molecular data suggest a rapid diversification event approximately 4 million years ago. These data indicate that the genus most closely related to

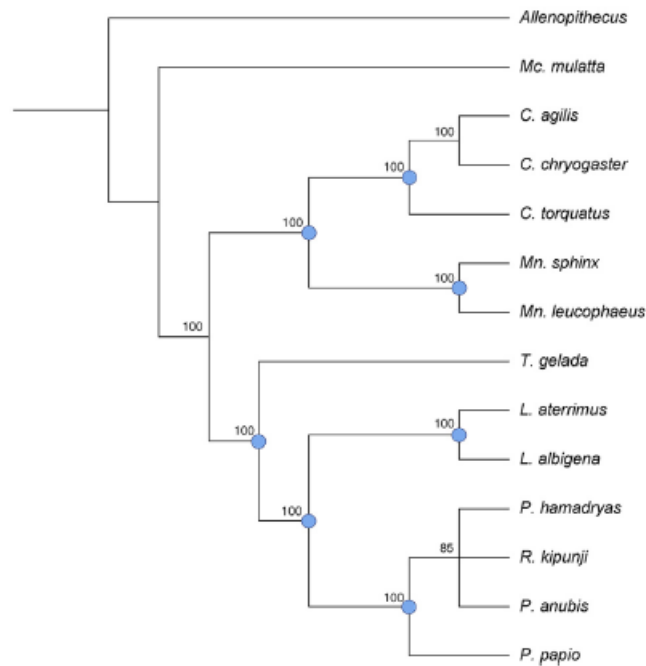


FIGURE 0.2 - Phylogeny of Papionini tribe. From Pugh and Gilbert 2018.

Papio is *Rungwecebus*, which is hypothesised to be a taxon resulting from hybridisation between *Papio* and *Lophocebus*, followed by *Lophocebus*, and ultimately *Theropithecus* (Pugh and Gilbert 2018; FIGURE 0.2).

The oldest fossil evidence attributable to the Papionini tribe is *Parapapio*, with remains discovered in East Africa dating from 7.4 to 5.3 million years ago (Leakey et al. 2003). This genus is documented in Southern Africa until about 2 million years ago (Delson, 1988).

Pliopapio alemui is another early Papionini fossil dated 5.2-4.2 million years ago and found in Ethiopia (Frost et al. 2009). His phylogenetic position remains uncertain. Frost (2001) proposed that he might be a stem Papionini, or (more likely) a basal member of the *Papio*/*Theropithecus*/*Lophocebus* clade.

Two other large-bodied Plio-Pleistocene Papionini are *Dinopithecus ingens* and *Gorgopithecus major*; they are mainly found in Southern Africa (Delson 1984), but they are not well understood. Freedman (1957) suggested a close relationship between *Dinopithecus* and *Gorgopithecus* based on similarities in their molar teeth. Both seem to be descendants of Southern African *Parapapio* (Jablonski 2002).

The *Lophocebus* genus fossil record is fragmentary or entirely absent (Jablonski and Frost 2010). In contrast, the genus *Theropithecus* is well-represented and widespread throughout the fossil record from the Plio-Pleistocene, with a lot of diverse forms documented (Jolly 1972; Jablonski 1993). The older known species were *T. darti*, *T. quadratiostris*, and *T. brumpti*; they were primarily folivorous. In contrast, the more recent species *T. oswaldi* likely had a diet that was more graminivorous (Benefit and McCrossin 1990). The decline and extinction of most *Theropithecus* species, with the exception of the surviving *T. gelada*, along with the extinction of *Parapapio*, appears to align with the rise and diversification of *Papio* species. This shift

suggests that the ecological niches previously occupied by the two genera are now filled by the newborn *Papio* genus (Jablonski 1993; 2002).

The oldest fossils assigned to the genus *Papio* all come from Southern Africa, most of them found at sites within the Krugersdorp "Cradle of Humankind" region, in South Africa. *P. izodi* was first described by Gear (1926), on material from the sites of Sterkfontein Members 2 and 4. Dated to between 3.7 and 2 Mya (Gilbert et al. 2015), it is not yet clear if it belongs to the genus *Papio*, if it is a generic member of the crown clade, or if it is a late-occurring stem taxon (Gilbert et al. 2018). A more secure member of the genus is *P. robinsoni*, described for the first time by Freedman (1957) and represented by fossil material found at various sites across South Africa with dates ranging from 2.6 to 1.5 Mya (Gilbert et al. 2015). Contemporaneous to *P. robinsoni* is *P. angusticeps*, the form most closely resembling modern *Papio* species. Placed within the genus *Papio* for the first time by Freedman (1957), it is associated with remains that have been found across South Africa, the oldest estimated to date back to 2.3-2.0 Mya (Gilbert et al. 2015). The only *Papio* species associated with remains found outside South Africa is *P. botswanae*, currently represented by a single specimen. Its description is based on the skull found in the layer of the Middle Pleistocene (between 200-400 thousand years) in the cave system of !Nemutsa Hill in Botswana (Williams et al. 2012). Findings assigned to *Papio* did not appear in Eastern Africa until the Middle Pleistocene (Jablonski and Leakey 2008)

On the basis of currently available fossil data, Elton and Dunn (2020) infer that the genus originated in the Southern tropic region of Africa (Jolly 2020), and from this area, multiple waves of dispersal have departed. The first wave was probably driven by *P. angusticeps*-like baboons to the South, leaving traces with the fossils we mentioned above. This first wave, according to Elton and Dunn (Elton and Dunn 2020), would have brought a primitive chacma-like form to the North-East of Namibia before dispersing in the East around the Kalahari Desert and in the Northern part of South Africa. According to Elton and Dunn (2020), after the split between the Northern and Southern *Papio* clades, another wave of dispersion towards the South may have occurred, this time by populations of the Northern clade, in particular of *P. robinsoni*, which shows its presence in South Africa with fossil remains, as described above. Finally, a second chacma clade originated later due to allopatric divergence due to the presence of barriers created by the aridity of that area; these baboons' lines dispersed towards the cape and subsequently towards Namibia. This would have led to a replacement in South Africa of the earliest forms of the genus *Papio* (Elton and Dunn 2020).

During the Middle Pleistocene, there were several speciation events which gave rise to different species. As regards the type of evolutionary model for these speciation events, there are two different hypotheses. The first supports allopatric speciation followed by contacts and gene flow, whereby populations were initially separated due to the emergence of barriers, most likely due to climate change and consequent habitat modification (Chala et al. 2019). The second hypothesis instead suggests a model of speciation of the parapatric type (Fuchs et al. 2018) for which the distribution range of the genus consists of a gradual change of environments so that by moving across this region the different populations have adapted to the new environment, diverging from the original source population and generating intermediate hybridisation zones between the newly formed species. Both of these hypotheses appear to be valid and well supported by the available data, so further studies are needed to find the most probable synthesis and model.

III.V - Baboons Today

Baboons are terrestrial quadrupedal animals with a muzzle resembling that of a dog (Fischer et al. 2019). The name “cynocephalus” derives from the Greek κύων (dog) and κεφαλή (head), meaning dog-headed; whereas the name “anubis” refers to the Egyptian God Anubis, depicted with the head of a jackal.

In these animals, body mass varies greatly among species (the smallest being Kinda, while the largest is chacma), but the marked sexual dimorphism remains consistent, with males having approximately twice the body mass of females (FIGURE 0.3; Fischer et al. 2019).

The six species differ significantly in terms of fur colour and sexual displays (FIGURE 0.3). Females of all species, upon reaching sexual maturity, exhibit sexual swellings in the anogenital region. These swellings change during the menstrual cycle to indicate the most suitable mating period for males (Fischer et al. 2019).

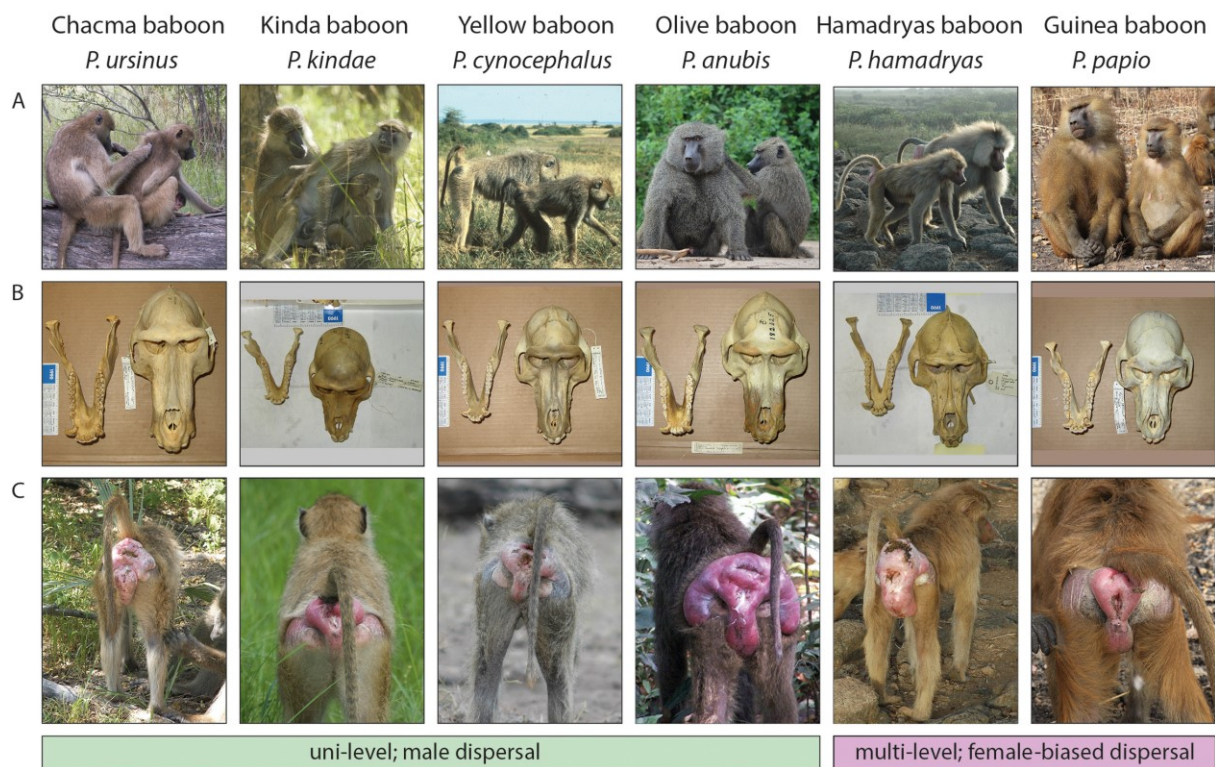


FIGURE 0.3 - Main traits of baboon species: A) phenotypic variation (males and females), B) skull variation in males, and C) sexual swelling in females during oestrus. Species are divided by social organisation and dispersal behaviour. Photos from Alexis Amann, Andrea Cardini, Sarah Elton, Julia Fischer, Courtney Fitzpatrick, James Higham, Megan Petersdorf, Joan Silk and Larissa Swedell. Figure from Fisher et al. 2019.

All species are largely terrestrial during daylight hours, retreating to trees or cliffs during night-time for greater protection from nocturnal predators. They are generalist animals, highly flexible both in terms of habitat (they inhabit semi-deserts, prairies, woodlands, savannas, humid forests, and Afro-alpine grasslands up to 3000 m above sea level) and diet (they feed on fruits, seeds, leaves, roots, as well as arthropods, small mammals, and birds; Fischer et al. 2019).

Like most primates, baboons are social animals, but behaviour and social organisation (including group size) vary by species. Chacma, yellow, and olive baboons have group sizes ranging from a dozen to several hundred individuals, while hamadryas and Guinea baboons only temporarily aggregate in groups of hundreds. Chacma, olive, Kinda, and yellow baboons have multi-male, multi-female societies where related females form the core of the group, and males leave their natal groups to join others, indicating male dispersal. In these societies, clear hierarchies exist among males and females based on the aggressiveness of their interactions. The primary social bond among female chacma, olive, and yellow baboons is kinship, while in Kinda baboons, friendships formed through grooming between males and females are also significant. Females in these four species interact and mate with multiple males within the group, showing no clear gregariousness, although higher-ranking males mate more frequently, though not exclusively, with females in the group (Fischer et al. 2019).

In contrast, hamadryas baboons live in multi-level societies with reproductive units consisting of a dominant male, a variable number of females, and sometimes several bachelor males. Reproductive units associate temporarily during the day, forming larger groups. The leading male and bachelor males tend to be matrilineally related, as they rarely leave their original group; thus, dispersal is predominantly female. A similar structure has been observed in Guinea baboons (Fischer et al. 2019).

Phylogenetically, baboons are divided into two major groups: Northern and Southern (Rogers et al. 2019). The Southern clade appears to be the oldest, as an exceptionally ancient mitochondrial lineage exclusive to the modern Southern chacma baboons is observed, representing the first divergence of approximately 2.45 [2.07–2.81] mya (Zinner et al. 2009; Roos et al. 2021). This haplotype has also been identified in the remains of an ancient individual of *P. ursinus* dating back 5,800 years, discovered in Lesotho, demonstrating a certain long-term continuity in the Southern region of South Africa (Mathieson et al. 2020). However, it is unclear whether the Southern chacma baboons were the first to diverge or if the ancestral baboon lineage initially split into major Southern and Northern clades before other divergences occurred. In fact, the split between the Northern and Southern clades is dated slightly more recently, but with an interval that fully overlaps (2.29 [1.95–2.65] mya (FIGURE 0.4; Zinner et al. 2009; Roos et al. 2021)).

Subsequently, between 2.2 and 1.6 mya, there was the divergence of yellow and Kinda baboon lineages. Additionally, around 0.8 million years ago, the Southern yellow baboons underwent a separation from the Northern chacma baboon lineage. Furthermore, in the context of the mitochondrial lineage of yellow baboons, it is noteworthy that more northerly populations exhibit shared haplogroups with hamadryas and olive baboons observed in East Africa. Olive baboons from central Africa exhibit haplogroup sisters to the yellow/hamadryas/Eastern olive clade; instead, the olive baboons of West Africa exhibit haplotypes that are sister to those of Guinea baboons. The olive baboons from Sierra Leone share haplogroups with Guinea baboons suggesting introgression, while the other Western olive baboons have haplogroups sisters to this clade (FIGURE 0.4; Zinner et al. 2009; Roos et al. 2021).

However, mitochondrial analyses indicate a much more complex evolutionary history with numerous haplogroups (ten major well-supported haplotypes; (Roos et al. 2021)) reflecting geographic origins better than morphological characteristics or taxonomic classification (Rogers et

al. 2019; Roos et al. 2021) and all those observations reveal widespread paraphyly and polyphyly across nearly all species, which can be explained by extensive and iterated introgression phenomena during the evolutionary history of baboons (FIGURE 0.4).

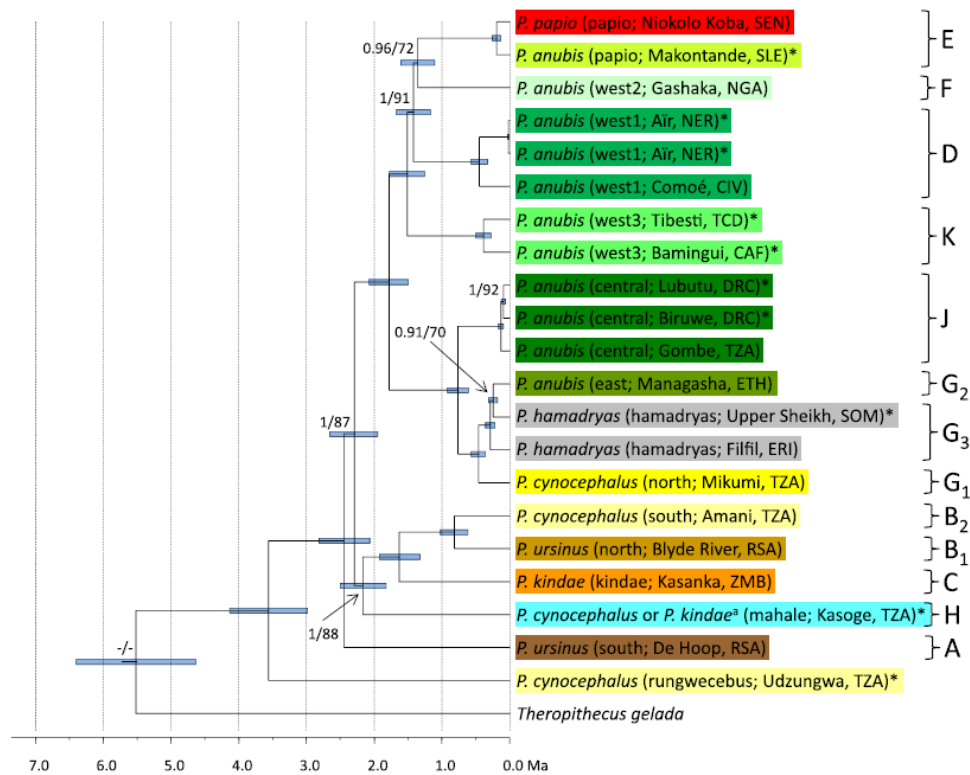


FIGURE 0.4 - Dated mitochondrial DNA phylogeny of the genus *Papio*. From Roos et al. 2021.

Analyses of the nuclear genome also confirm the complex evolutionary history of these species, which includes ancient introgression events (Rogers et al. 2019; Santander et al. 2022; Sørensen et al. 2023). Multiple episodes of recent introgressions are also evident in the genomic composition of modern baboons. Many hybridisation zones have been identified: Awash region in Ethiopia between olive and hamadryas baboons, Amboseli in Kenya between olive and yellow baboons, Mozambique between yellow and chacma baboons, and Zambia between chacma and Kinda baboons (Cortés-Ortiz 2017; Zinner, Arnold, and Roos 2011). More details on ancient and recent hybridisations will be provided in the following chapters.

In addition to intragenetic hybridisation, baboons are known to hybridise with closely related genera. A particularly intriguing case involves intergeneric hybridisation in Tanzania approximately 0.35 million years ago between yellow baboons and the kipunji (*Rungwecebus kipunji*). This species was discovered in two isolated populations in the Southern highlands of Tanzania (Davenport et al. 2006). Following the identification of this new taxon, genetic analysis revealed that the first studied individual possessed a mtDNA closely related to yellow baboons from the Southern range, which live parapatrically or sympatrically with the kipunji (Burrell et al. 2009). Phylogenetic analysis indicated that there was introgression of maternally inherited mitogenome from baboons to *Rungwecebus*. Another case of intergeneric introgression, which has been observed as potentially ongoing, involves hybridisation between baboons and *Theropithecus gelada*, as previously mentioned and further discussed in Chapter 3 of this work.

IV - Aim of the Work

This work aims to study the evolutionary history of baboons through their genomes, with a particular focus on investigating the events of hybridisation that have occurred between different species over time and how these have shaped population structure within species. Specifically, the first chapter will focus on the population structure of *Papio ursinus* and the hybridisation with the two species, *Papio cynocephalus* and *Papio kindae*, Whose ranges border that of the chacma baboon to the North. In the second chapter, museum data will be used to supplement the limited dataset available for two *Papio* species in East Africa (*Papio anubis* and *Papio hamadryas*), where their ranges overlap, particularly in Ethiopia. The population structures of the two species and signs of hybridisation in different populations will be investigated. Finally, in the third chapter, the aim will be to understand the evolutionary history of *Theropithecus gelada* by studying its population structure and population dynamics including hybridisation with baboons in East Africa.

V - References

- Abbott, R. J., Barton, N. H., and Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular ecology*, 25(11), 2325–2332. <https://doi.org/10.1111/mec.13685>
- Aldhebiani A. Y. (2018). Species concept and speciation. *Saudi journal of biological sciences*, 25(3), 437–440. <https://doi.org/10.1016/j.sjbs.2017.04.013>
- Arnold, M. L., and Meyer, A. (2006). Natural hybridization in primates: one evolutionary mechanism. *Zoology (Jena, Germany)*, 109(4), 261–276. <https://doi.org/10.1016/j.zool.2006.03.006>
- Benefit, B. R., and McCrossin, M. L. (1990). Diet, species diversity and distribution of African fossil baboons. *Kroeber Anthropological Society Papers*, 71-72, 77-93.
- Burrell, A. S., Jolly, C. J., Tosi, A. J., and Disotell, T. R. (2009). Mitochondrial evidence for the hybrid origin of the kipunji, *Rungwecebus kipunji* (Primates: Papionini). *Molecular phylogenetics and evolution*, 51(2), 340–348. <https://doi.org/10.1016/j.ympev.2009.02.004>
- Chala, D., Roos, C., Svenning, J. C. and Zinner, D. (2019). Species-specific effects of climate change on the distribution of suitable baboon habitats - Ecological niche modeling of current and Last Glacial Maximum conditions. *Journal of human evolution*, 132, 215–226. <https://doi.org/10.1016/j.jhevol.2019.05.003>
- Cortés-Ortiz, L. (2017). Hybridization and Hybrid Zones. In *The International Encyclopedia of Primatology*, 1–5. John Wiley & Sons, Inc, Hoboken, NJ, USA. <https://doi.org/10.1002/9781119179313.wbprim0380>.

Davenport, T. R., Stanley, W. T., Sargis, E. J., De Luca, D. W., Mpunga, N. E., Machaga, S. J., and Olson, L. E. (2006). A new genus of African monkey, *Rungwecebus*: morphology, ecology, and molecular phylogenetics. *Science (New York, N.Y.)*, 312(5778), 1378–1381.
<https://doi.org/10.1126/science.1125631>

Delson, E. (1988). Chronology of South African australopith site units. In Grine, F. (Ed.), *Evolutionary History of the “Robust” Australopithecines* (pp. 317–324). *Aldine de Gruyter*, New York.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. *Columbia University Press*.

Dominy, N. J., Ikram, S., Moritz, G. L., Wheatley, P. V., Christensen, J. N., Chipman, J. W., and Koch, P. L. (2020). Mummified baboons reveal the far reach of early Egyptian mariners. *eLife*, 9, e60860. <https://doi.org/10.7554/eLife.60860>

Dunbar, R. I. M., and Patsy Dunbar. (1974). On Hybridization between *Theropithecus gelada* and *Papio anubis* in the Wild. *Journal of Human Evolution* 3 (3): 187–92.

Eldredge, N. (1993). What, If Anything, Is a Species? In *Species, Species Concepts and Primate Evolution*, 3–20. *Springer US*, Boston.

Elton, S., and Dunn, J. (2020). Baboon biogeography, divergence, and evolution: Morphological and paleoecological perspectives. *Journal of human evolution*, 145, 102799.
<https://doi.org/10.1016/j.jhevol.2020.102799>

Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., Seyfarth, R. M., ... Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. *eLife*, 8, e50989. <https://doi.org/10.7554/eLife.50989>

Fitzpatrick, B. M., Ryan, M. E., Johnson, J. R., Corush, J., and Carter, E. T. (2015). Hybridization and the Species Problem in Conservation. *Current Zoology* 61 (1): 206–16.

Freedman, L. (1957). Fossil Cercopithecoidea of South Africa. *Annals of the Transvaal Museum*, 23, 121–262.

Frost, S. R., Haile-Selassie, Y., and Hlusko, L. G. (2009). Cercopithecidae. In Haile-Selassie, Y., WoldeGabriel, G. (Eds.), *Ardipithecus Kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia* (pp. 135–158). *University of California Press*, Berkeley.

Fuchs, A. J., Gilbert, C. C., and Kamilar, J. M. (2018). Ecological niche modeling of the genus *Papio*. *American journal of physical anthropology*, 166(4), 812–823.
<https://doi.org/10.1002/ajpa.23470>

Futuyma, D. and Kirkpatrick, M. (2017). *Evolution*. *Sinauer*.

Gear, J. H. S. (1926). A Preliminary Account of the Baboon Remains from Taungs. *South African Journal of Science XXIII*.

Gilbert, C. C., Frost, S. R., Pugh, K. D., Anderson, M., and Delson, E. (2018). Evolution of the modern baboon (*Papio hamadryas*): A reassessment of the African Plio-Pleistocene record. *Journal of human evolution*, 122, 38–69. <https://doi.org/10.1016/j.jhevol.2018.04.012>

Gilbert, C. C., Steininger, C. M., Kibii, J. M., and Berger, L. R. (2015). *Papio* Cranium from the Hominin-Bearing Site of Malapa: Implications for the Evolution of Modern Baboon Cranial Morphology and South African Plio-Pleistocene Biochronology. *PloS one*, 10(8), e0133361. <https://doi.org/10.1371/journal.pone.0133361>

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., ... Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, 328(5979), 710–722. <https://doi.org/10.1126/science.1188021>

Groves, C. P. (2001). *Primate Taxonomy*. *Smithsonian Institution Press*, Washington, DC.

Grubb, P., Butynski, T. M., Oates, J. F., Bearder, S. K., Disotell, T. R., Groves, C. P., & Struhsaker, T. T. (2003). Assessment of the diversity of African Primates. *International Journal of Primatology*, 24, 1301–1357.

Harrison, R. G., and Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular ecology*, 25(11), 2454–2466. <https://doi.org/10.1111/mec.13582>

Hamilton III, William, J. (1986). Namib Desert chacma baboon (*Papio ursinus*) use of food and water resources during a food shortage. *Madoqua*, 1986(4), 397–407.

Hey, J. (2001). *Genes, Categories, and Species: The Evolutionary and Cognitive Cause of the Species Problem*. *Oxford University Press*.

Kitchener, A. C., Machado, F. A., Hayssen, V., Moehlman, P. D., & Viranta, S. (2020). Consequences of the misidentification of museum specimens: the taxonomic status of *Canis lupaster soudanicus*. *Journal of Mammalogy*, 101(4), 1148–1150.

Jablonski, N. G. (Ed.). (1993). *Theropithecus: The Rise and Fall of a Primate Genus*. *Cambridge University Press*, Cambridge, UK.

Jablonski, N. G. (2002). Fossil Old World monkeys: the late Neogene radiation. In Hartwig, W.C. (Ed.), *The Primate Fossil Record* (pp. 255–299). *Cambridge University Press*, Cambridge, UK.

Jablonski, N. G., and Leakey, M. G. (2008). The importance of the Cercopithecoidea from the Koobi Fora Formation in the context of primate and mammalian evolution. In Jablonski, N.G., Leakey, M.G. (Eds.), *Koobi Fora Research Project: Volume 6: The Fossil Monkeys* (pp. 397–416). *California Academy of Sciences*, San Francisco.

Jablonski, N. G., and Frost, S. R. (2010). Cercopithecoidea. In Werdelin, L., Sanders, W.J. (Eds.), *Cenozoic Mammals of Africa* (pp. 393–428). *University of California Press*, Berkeley, CA.

Jolly, C. J. (1972). The classification and natural history of *Theropithecus* (*Simopithecus*) (Andrews, 1916) baboons of the African Plio-Pleistocene. *Bulletin of the British Museum of Natural History*, 22, 1–123.

Jolly, C. J. (1993). Species, Subspecies, and Baboon Systematics. In *Species, Species Concepts and Primate Evolution*, 67–108. *Springer US*, Boston.

Jolly C. J. (2001). A proper study for mankind: Analogies from the Papionin monkeys and their implications for human evolution. *American journal of physical anthropology, Suppl 33*, 177–204. <https://doi.org/10.1002/ajpa.10021>

Jolly C. J. (2020). Philopatry at the frontier: A demographically driven scenario for the evolution of multilevel societies in baboons (*Papio*). *Journal of human evolution*, 146, 102819. <https://doi.org/10.1016/j.jhevol.2020.102819>

Kanthaswamy, S., Ng, J., Satkoski Trask, J., George, D. A., Kou, A. J., Hoffman, L. N., Doherty, T. B., Houghton, P., and Smith, D. G. (2013). The genetic composition of populations of cynomolgus macaques (*Macaca fascicularis*) used in biomedical research. *Journal of medical primatology*, 42(3), 120–131. <https://doi.org/10.1111/jmp.12043>

Leakey, M. G., Teaford, M. F., and Ward, C. V. (2003). Cercopithecidae from Lothagam. In Leakey, M.G., Harris, J.M. (Eds.), *Lothagam: The Dawn of Humanity in Eastern Africa* (pp. 201–248). *Columbia University Press*, New York.

Liedigk, R., Roos, C., Brameier, M., and Zinner, D. (2014). Mitogenomics of the Old World monkey tribe Papionini. *BMC evolutionary biology*, 14, 176. <https://doi.org/10.1186/s12862-014-0176-1>

Mathieson, I., Abascal, F., Vinner, L., Skoglund, P., Pomilla, C., Mitchell, P., Arthur, C., Gurdasani, D., Willerslev, E., Sandhu, M. S., and Dewar, G. (2020). An Ancient Baboon Genome Demonstrates Long-Term Population Continuity in Southern Africa. *Genome biology and evolution*, 12(4), 407–412. <https://doi.org/10.1093/gbe/evaa019>

Mayr, E. (1942). *Systematics and the Origin of Species*. *Columbia University Press*, New York.

Nater, A., Mattle-Greminger, M. P., Nurcahyo, A., Nowak, M. G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T. B., Roos, C., Lameira, A. R., Wich, S. A., Askew, J., Davila-

Ross, M., Fredriksson, G., de Valles, G., Casals, F., Prado-Martinez, J., Goossens, B., Verschoor, E. J., ... Krützen, M. (2017). Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current biology: CB*, 27(22), 3487–3498.e10. <https://doi.org/10.1016/j.cub.2017.09.047>.

Phillips-Conroy, J. E., & Jolly, C. J. (1986). Changes in the structure of the baboon hybrid zone in the Awash National Park, Ethiopia. *American Journal of Physical Anthropology*, 71, 337–350.

Pugh, K. D., and Gilbert, C. C. (2018). Phylogenetic relationships of living and fossil African papionins: Combined evidence from morphology and molecules. *Journal of human evolution*, 123, 35–51. <https://doi.org/10.1016/j.jhevol.2018.06.002>

de Queiroz, K. (2011). Branches in the Lines of Descent: Charles Darwin and the Evolution of the Species Concept. *Biological Journal of the Linnean Society*. Linnean Society of London 103 (1): 19–35.

Regan, C. T. (1926). Organic evolution. *Report of the British Association for the Advancement of Science, 93rd Meeting*, 75-86.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., ... Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), 1053–1060. <https://doi.org/10.1038/nature09710>

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M., Ko, Y. C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S., & Stoneking, M. (2011). Denisova Admixture and the first modern human dispersals into Southeast Asia and Oceania. *American journal of human genetics*, 89(4), 516–528. <https://doi.org/10.1016/j.ajhg.2011.09.005>

Reich, D. (2018). Who We are and how We Got Here: Ancient DNA and the New Science of the Human Past. *Oxford University Press*.

Richard, A. (1985). Primates in Nature. *V. H Freeman*, San Francisco.

Rogers, J., Raveendran, M., Harris, R. A., Mailund, T., Leppälä, K., Athanasiadis, G., Schierup, M. H., Cheng, J., Munch, K., Walker, J. A., Konkel, M. K., Jordan, V., Steely, C. J., Beckstrom, T. O., Bergey, C., Burrell, A., Schrempf, D., Noll, A., Kothe, M., Kopp, G. H., ... Baboon Genome Analysis Consortium (2019). The comparative genomics and complex population history of *Papio* baboons. *Science advances*, 5(1), eaau6947. <https://doi.org/10.1126/sciadv.aau6947>

Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American journal of physical anthropology*, 174(3), 407–417.

<https://doi.org/10.1002/ajpa.24186>

Santander, C., Molinaro, L., Mutti, G., Martínez, F. I., Mathe, J., Ferreira da Silva, M. J., Caldon, M., Oteo-Garcia, G., Aldeias, V., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Hammond, P., Lüdecke, T., Pinto, M. J., Meira Paulo, L., Stalmans, M., Regala, F. T., ... Capelli, C. (2022). Genomic variation in baboons from central Mozambique unveils complex evolutionary relationships with other *Papio* species. *BMC ecology and evolution*, 22(1), 44. <https://doi.org/10.1186/s12862-022-01999-7>

Smuts, B. B., Seyfarth, R. M., Wrangham, R. W., and Struhsaker, T. T. (eds.), 1986. Primate Societies, pp. 121-135. *University Press*, Chicago.

Sober, E., and Orzack, S. H. (2003). Common Ancestry and Natural Selection. *The British Journal for the Philosophy of Science* 54 (3): 423–37.

Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., Storer, J. M., Kuhlwilm, M., Fontserè, C., Seshadri, L., Bergey, C. M., Burrell, A. S., Bergman, J., Phillips-Conroy, J. E., Shiferaw, F., Chiou, K. L., Chuma, I. S., Keyyu, J. D., Fischer, J., Gingras, M. C., ... Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science (New York, N.Y.)*, 380(6648), eabn8153. <https://doi.org/10.1126/science.abn8153>

Stringer, C. B., and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, 239, 1263–1268.

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

VandeBerg, J. L., Williams-Blangero, S., and Tardif, S. D. (2009). The Baboon in Biomedical Research. *Springer*, New York.

Van Neer, W., Udrescu, M., Peters, J., De Cupere, B., Pasquali, S., and Porcier, S. (2023). Palaeopathological and demographic data reveal conditions of keeping of the ancient baboons at Gabbanat el-Qurud (Thebes, Egypt). *PloS one*, 18(12), e0294934. <https://doi.org/10.1371/journal.pone.0294934>

Van Valen, L. (1976). Ecological Species. Multispecies, and Oaks. *Taxon* 25, 233–239.

Wiley, E. O. (1981). Remarks on Willis' Species Concept. *Systematic Biology* 30 (1): 86–87.

Wilkins, John Simpson (2003). The Origins of Species Concepts. Dissertation, *University of Melbourne*.

Williams, B. A., Ross, C. F., Frost, S. R., Waddle, D. M., Gabadirwe, M., & Brook, G. A. (2012). Fossil *Papio* cranium from !Ncumtsa (Koanaka) Hills, Western Ngamiland, Botswana. *American*

Journal of Physical Anthropology, 149, 1-17.

Wolpoff, M. H., Wu, X., and Thorne, A. G. (1984). Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia. In Smith, F.H., Spencer, F. (Eds.), *The Origins of Modern Humans: A World Survey of the Fossil Evidence* (pp. 411–483). New York: *Alan R. Liss*.

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, 83. <https://doi.org/10.1186/1471-2148-9-83>

Zinner, D., Arnold, M. L., and Roos, C. (2009). Is the new primate genus *rungwecebus* a baboon?. *PloS one*, 4(3), e4859. <https://doi.org/10.1371/journal.pone.0004859>

Zinner, D., Arnold, M. L., and Roos, C. (2011). The strange blood: natural hybridization in primates. *Evolutionary anthropology*, 20(3), 96–103. <https://doi.org/10.1002/evan.20301>

Zinner, D., Buba, U., Nash, S., Roos, C. (2011). Pan-African Voyagers: The Phylogeography of Baboons. In: Sommer, V., Ross, C. (eds) *Primates of Gashaka. Developments in Primatology: Progress and Prospects*, vol 35. *Springer*, New York, NY.

Introduction - Overview of the Materials and Methods

In this chapter, a general overview of the materials and methods utilized in this work will be provided. Detailed descriptions and specifications will be outlined in the relevant sections within each of the three chapters. This structure aims to provide a general background before delving into the precise methodologies and materials used, ensuring clarity and comprehensiveness throughout the thesis.

First and foremost, it is essential to acknowledge the High-Performance Computing facility of the University of Parma (HPC.unipr.it), which hosted all the analyses conducted during this work. However, it is equally important to highlight some of the critical issues that this infrastructure presents, as they had a significant impact on the progression of the research activities. Working with medium-to-high coverage genomes requires a substantially extended storage capacity. The limited storage capacity allocated to our research group (and individual users) proved insufficient to fully meet the project's needs. This limitation inevitably slowed down and hindered the progress of the work, necessitating reductions in the original dataset and the deletion of newly produced material that was not immediately needed. Consequently, many files had to be created multiple times (in the same manner), as they were found to be useful long after their initial creation and subsequent deletion.

Despite these challenges, it is important to emphasize that the HPC infrastructure still provided crucial support for completing the analyses. In the future, an enhancement of the available storage resources would be desirable to facilitate research on large-scale datasets and improve the overall efficiency of the analyses.

VI - Dataset preparation

In this study, we utilized whole genomic data derived from individuals in their natural habitat, those born or raised in captivity (zoos), or from individuals who had been deceased for varying lengths of time before tissue sampling for DNA extraction. These samples included those from ancient times as well as those from the relatively recent past (XIX-XX century), which had been processed and preserved in museums. The data employed in this study were predominantly sourced from previously published works. However, in some cases, novel genomic data were generated and published concurrently with the studies described. This includes all complete genomes of museum specimens of baboons, as well as individuals in the wild and/or in captivity of geladas, chacma baboons, and *Lophocebus aterrimus*. DNA from these samples was extracted either by external facilities or through collaborations with other universities (details will be provided in the introductions of the respective studies).

The raw sequences were mapped to a reference genome. Mapping the reads to a reference genome involves aligning the short sequences obtained from DNA extraction and sequencing to a pre-annotated and assembled genome, thereby producing a continuous sequence from many thousands of short reads. A critical aspect of this process is the sequencing coverage (i.e., the proportion of the reference genome that can be covered by the obtained reads), and the depth (i.e., the average number of reads aligning to a given position). The choice of reference genome is crucial for accurate data analysis. When aligning the short reads obtained through sequencing, the assembled genome tends to be homologous to the reference genome used. Therefore, studying hybridisations between different species and mapping everything onto the reference genome of one of the two species risks invalidating the analyses, because the second species appears much more similar to the first than it is. On the other hand, using a reference genome from a species too distantly related to the species under consideration risks overlooking a substantial amount of variability. This is because the divergence points between the species to which the reference genome belongs and the considered species is too ancient: relative to that moment, all variants appear derived.

There are several methods for aligning sequences to a reference genome, and in this study, the *bwa-mem2* software (v.2.2.1; Vasimuddin et al. 2019) was employed. The final output is a BAM file (file.bam), a compressed binary format file that includes a header containing information such as the file name, size, and alignment method, and an alignment section containing the read sequences, their quality, alignment information, and read names along with chromosome, coordinates, alignment quality, and various other details. The BAM file is always accompanied by a BAM index file (file.bam.bai) which provides the index for the corresponding BAM file.

The content of the BAM files is then processed by marking duplicate sequences and filtering based on sequencing quality using specific tools. Once the final files are obtained, variant calling is performed using the *GATK* software (v. 4.2.4.1; McKenna et al. 2010), generating a new file, the gVCF, which contains all the allelic variants present in the individual's genome. The individual gVCFs are then merged into a single VCF file containing the variants of all considered individuals. Both gVCF and VCF files consist of a header containing information about the file format version, the variant caller, the reference genome used, and the BAM file from which these files were generated. The last line of the header contains the column names for the subsequent data section. In this second section, each row contains information about individual variants such as the chromosome and its position, the variant name, the reference and alternative alleles, the variant quality score, genotype, and other indices. The gVCF file differs from the VCF in that it contains all genome sites, whether a variant is called or not. Thus, besides merging the various gVCFs into a single file, converting to VCF allows retaining only those sites that are truly variable.

The final VCF file undergoes hard filtering, which involves setting specific thresholds for one or more annotations and discarding all variants with annotation values above or below the set thresholds. For subsequent analyses, the VCF file of the variants was converted into the PLINK BED and EIGENSTRAT EIGENSOFT formats. Both formats consist of three separate files, each containing different types of information.

The BED format includes a .bed file, which represents genotypes in a binary format. The first three bytes are standard and confirm that it is indeed a .bed file; subsequent bytes provide

information on four genotypes each, indicating whether SNPs are homozygous for the first allele (00) or the second (11), heterozygous (10), or genotype missing (01). The information to interpret this file is provided in the second file, the .bim, which specifies the chromosome where the SNP is located, its name, its centiMorgan position, physical coordinates, and the nucleotides of the first (minor allele frequency) and second alleles (major allele frequency). The third file is the .fam, providing information on the group identifier each individual belongs to, the individual's ID, the parents' IDs, sex, and phenotype.

The EIGENSTRAT format comprises a .geno file containing genotype information organized so that each line represents a variant, where individuals are represented by a single character: 2 (homozygous for the reference allele), 1 (heterozygous), 0 (homozygous for the alternate allele), and 9 (SNP missing for that individual). The second file is the .snp, containing all information about each variant; it also consists of one line per SNP with six columns specifying the SNP name, chromosome, Morgans position, physical base position, and the two allelic variants. The third file is the .ind, containing one line per individual with information on the individual's identifier code, sex, and population.

Regarding museum data, the approach was slightly different due to the highly degraded sequences, low coverage, and hence, many SNPs present in genomes with medium to high coverage are absent. These sequences, once mapped to the reference genome, were filtered with much more lenient quality parameters. From the BAM, the process moved directly to the final format without passing through gVCF, as SNP calling was conducted using a specialized tool for ancient DNA (*PileupCaller*; Lamnidis et al. 2018), which calls variants based on the SNP list provided by the VCF of modern genomes with medium to high coverage. The output of this software is directly a triplet of EIGENSTRAT files. The same procedure was followed for modern genomes with low coverage.

VII - Nuclear Genome Analysis

Once the final files were obtained, genomic data analysis was conducted. The methods used for the analyses in the following three chapters will be illustrated below. Specific analyses conducted in a single work and the operative details for all the methods will be detailed in the relevant sections.

VII.I - Principal Component Analysis

Principal component analysis (PCA) is a type of multivariate statistical analysis that involves combining the original variables to maximize the variance of individual components (PCA1, PCA2,...,PCAn). This process generates new variables, each representing the majority of the variability present among the original ones. This allows for the representation of the variability among individuals using a reduced number of components that describe a high percentage of the overall variance. The points corresponding to the different individuals analysed are positioned farther apart the more different they are from each other.

These analyses were conducted using the EIGENSOFT *smartpca* software v.8.0.0 (Patterson, Price, and Reich 2006), a method specifically designed for studying population structure using

genomic data. *smartpca* requires genomic data in EIGENSTRAT format as input and provides two distinct output files: one in .evec format showing the principal components (the eigenvectors), and the other in .eval format providing a list of all the eigenvalues, which are the scaling factors for the eigenvectors. The ratio of a single eigenvalue k to the sum of all eigenvalues represents the proportion of variance explained by eigenvector k .

Compared to other software that performs similar analyses, *smartpca* is faster with large datasets and does not attempt to classify all individuals into distinct populations or linear combinations of populations, which might not always be the correct model. PCA was conducted exclusively with modern genomes of medium to high coverage, as those with low coverage and museum samples have a significantly reduced number of variants on which to perform the analysis. However, these low-coverage and ancient samples were projected onto the obtained PCA. Projecting ancient samples onto modern ones was necessary because ancient genotypes often have substantial differences in SNP coverage, which would otherwise distort the final PCA.

VII.II - D-Statistics or ABBA-BABA test

This is a statistical method to test the occurrence of any imbalance in allele sharing between four taxa (((W,X),Y),Z). The test considers the ancestral alleles (A) and the derived alleles (B) in the genomes of the four taxa (FIGURE 0.5; Patterson et al. 2012). The estimation of D is based on allele frequencies and is calculated for each SNP as follows:

$$D = \frac{\sum (w - x)(y - z)}{\sum (w + x - 2wx)(y + z - 2yz)}$$

Along with the value of D , the software also generates the Z-score, which is the number of standard errors by which D deviates from 0. An absolute Z-score value greater than 3 is generally accepted as significant.

If there has been no gene flow, meaning W and X share the same amount of alleles with Y , then the value of D will be zero, or at least the Z-score will be between 3 and -3. When there is gene flow, there is an excess of one of the two allelic patterns, which produces a deviation of D from zero: a positive D and $Z > 3$ (excess of BABA) indicates introgression between W and Y , whereas a negative D and $Z < -3$ (excess of ABBA) points to gene flow between X and Y .

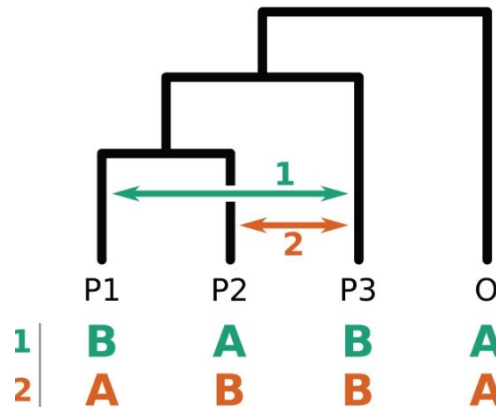


FIGURE 0.5 – D-statistics. Introgression events can result in a significant excess of ABBA or BABA patterns according to the D-statistic in a phylogeny where P1 (W), P2 (X), P3 (Y) and O (Z) are different taxa. From modified from Tricou et al. 2022 (<https://doi.org/10.1093/sysbio/syac011>).

In this analysis, Z is the outgroup. The choice of the outgroup is crucial as it helps to polarize the variants as ancestral or derived. Therefore, it must be sufficiently distant from all the taxa involved to ensure it is not involved in any hybridisation with them, but not excessively distant, otherwise, artifacts may be created. Here, *Mandrillus leucophaeus*, or drill, was chosen because it is not too distant from all the considered species, yet external to the *Papio-Lophocebus-Theropithecus* group. Additionally, its geographic location, far from all the considered species (West Africa), makes it reliably distant from any gene flow that may have occurred in recent evolutionary times.

VII.III - f4-ratio

The f4-ratio is a method used to estimate the proportion of ancestry in an admixed population under the assumption that we have a correct historical model of the hybridisation event (Patterson et al. 2012). To perform this analysis, five taxa are necessary: A, B, X, C, and O. Taxon X is the taxon for which we want to estimate the proportion of ancestry following a hypothetical hybridisation event. B and C are the sources of introgression we want to evaluate; taxon A is the sister group of B, and O is the outgroup (FIGURE 0.6). The proportion of admixture (alpha) of B in X is calculated using the ratio of two f4 tests (a method very similar to the D-statistics):

$$\hat{\alpha} = \frac{f_4(A, O; X, C)}{f_4(A, O; B, C)}$$

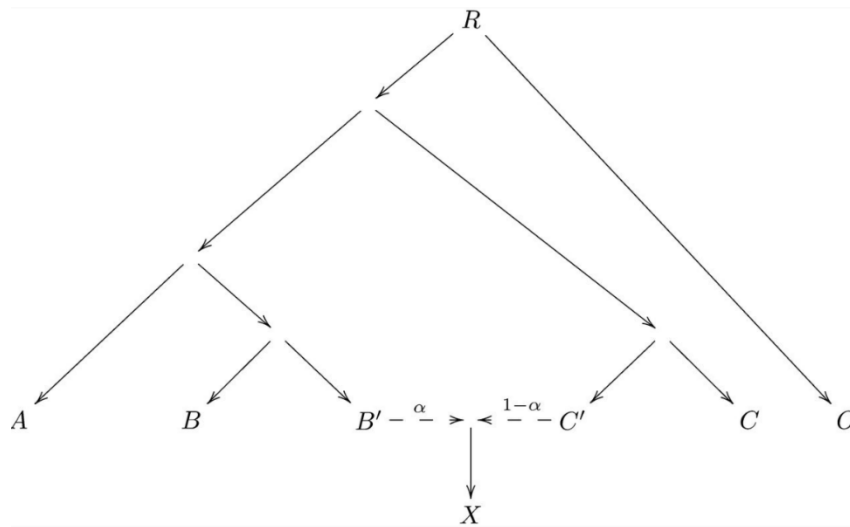


FIGURE 0.6 - f4-ratio estimation. From Patterson et al. 2012.

VII.IV - ADMIXTURE

One approach to studying populations and their origins is through the software ADMIXTURE (v.1.3.0; Alexander, Novembre, and Lange 2009), which allows for estimating the genomic contributions that ancestral populations have left in the populations under examination. This method involves model-based ancestry estimation, effectively estimating ancestry coefficients as parameters of a statistical model. Another program using this approach is STRUCTURE. Both programs simultaneously estimate population allele frequencies and the proportions of each ancestry, using them to model the probability of the observed genotypes. However,

ADMIXTURE focuses on maximizing the likelihood rather than posterior sampling, which is typical of STRUCTURE.

ADMIXTURE requires input consisting of genotypes from I unrelated individuals, with J SNPs, and composed of contributions from K ancestral populations. Each ancestral population k provides a fraction q_{ik} of an individual's genome. Therefore, it is crucial to know the number of ancestral populations K to input into the software for computation. Upon completion, ADMIXTURE provides two output files: a .Q file with a matrix indicating the fractions of ancestries and a .P file containing the inferred allele frequencies of the ancestral populations. In both cases, the number of rows corresponds to the number of samples, and the number of columns corresponds to the number of K .

VII.V - Pairwise Sequentially Markovian Coalescent (PSMC)

The Pairwise Sequentially Markovian Coalescent (PSMC) model is a statistical software tool that utilizes genome coalescence to reconstruct the population history of an individual by estimating the effective population size over time (Li and Durbin 2011). Genome coalescence is a concept that describes the process of tracing back the copies of a gene through time to their most recent common ancestor. In other words, it is a model that allows for the reconstruction of the genealogical history of a sample of genetic sequences, determining when and how these sequences originated from a single common ancestor.

The coalescent model assumes that genetic variations (mutations) accumulate over time and that the probability of coalescence (i.e., the point at which two genealogical lines meet at a common ancestor) depends on the population size and the structure of the population itself. Thus, this software retraces the coalescence points and estimates demographic fluctuations over time. The effective population size is calculated as a result of fluctuations in the size of a given population, but it also reflects changes in the degree of connectivity (e.g., migrations) within a structured population.

VII.VI - hPSMC

This is a variant of the methodology described in the previous paragraph, but it is useful for estimating the most recent genetic contact between two taxa (Cahill et al. 2016). In this case, it is necessary to construct an artificial hybrid using the X chromosomes of two males. It is important to use the X chromosome of a male because this sequence is haploid, ensuring there is no mixture of the two parental sequences. As known, males possess only one X chromosome. The Y chromosome is not used because it is much smaller and less stable. Once the hybrid is generated, the procedure is similar to PSMC. Here, genomic coalescence identifies the period of the most recent contact between the two artificial hybrid parents. As this approach nears the point of last contact, the number of coalescence points decreases significantly (indicating theoretically a larger population size in the past), tending towards infinity.

VII.VII - Heterozygosity and Run of Homozygosity

Heterozygosity is a measure of genetic variability within a population. It refers to the proportion of heterozygous genotypes among all genotypes in a population. This parameter ranges from 0 to 1, where the extremes represent the absence of heterozygosity and equal allele frequencies,

respectively. Populations with high heterozygosity tend to exhibit greater genetic variability, which can confer advantages in terms of adaptability and resistance to diseases.

Runs of Homozygosity (ROH) are stretches of the genome where an individual has two identical copies of the same DNA segment, inherited from both parents. These regions indicate that the individual has inherited the same alleles from both parents, suggesting some degree of consanguinity or recent common ancestry. ROHs form when an individual inherits identical DNA blocks from both parents, often due to reproduction between closely related individuals. This occurs because parents share a recent common ancestor, increasing the likelihood of transmitting identical DNA segments to their offspring. Longer ROH typically indicates more recent consanguinity, whereas shorter ROH may indicate shared ancestry further back in time.

VIII - References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655–1664.

<https://doi.org/10.1101/gr.094052.109>

Cahill, J. A., Soares, A. E., Green, R. E., and Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1699), 20150138. <https://doi.org/10.1098/rstb.2015.0138>

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Kharitanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., and Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature communications*, 9(1), 5018.

<https://doi.org/10.1038/s41467-018-07483-5>

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, 2019, pp. 314-324.
<https://doi.org/10.1109/IPDPS.2019.00041>

Chapter 1 - Population Structure and Hybridisations in *P. ursinus*

The writing of this chapter has benefited from three academic articles on the topic of hybridisation among baboon species in Southern Africa. One of these papers was published in 2022, while another was under review, and a third was in preparation at the time of writing this chapter:

- Santander et al. 2022: “Genomic variation in baboons from central Mozambique unveils complex evolutionary relationships with other *Papio* species”, published in BMC Ecology and Evolution.
- Caldon et al. under review: “Baboons at a crossroads: dissecting central Mozambique's genomic links with *Papio* neighbours”.
- Caldon et al. in preparation. “The impact of hybridisation on the population structure of baboon southernmost species *P. ursinus*”.

Preliminary results from this chapter were presented as an oral communication at the AAI (Associazione Antropologi Italiani) congress in September 2022 at Chieti, Italy.

1.1 - Introduction

Papio ursinus commonly referred to as chacma baboons (FIGURE 1.1), is the southernmost species of the genus *Papio* (under PSC). This species has been observed in the region between the Zambezi River and the Cape of Good Hope, with greater frequency in the well-watered East, but present also in the driest Western areas, mainly in river valleys and around oases (Jolly 1993). On the basis of morphological traits, three geographical variants are currently described in this species (FIGURE 1.2A; Jolly 1993): the ‘Typical chacma’ or ‘Cape chacma’ (recognised in the South and East *P. ursinus* range), the ‘Ruacana’ or ‘Kalahari chacma’ (present in the Western arid areas of Namibia and Angola), and the ‘Griseipes’ or ‘Grey-foot chacma’ (observed in the Northern part of *P. ursinus* range). A gradient of forms has been suggested to be present from South to North, where larger and darker forms are present in the South and animals in the North present some features more similar to the neighbouring species of yellow baboon (Jolly 1993; Martinez et al. 2019).



FIGURE 1.1 – Chacma baboon from Kruger National Park, South Africa. Photo by Bernard Dupont. Image via Flickr licensed under Creative Commons CC BY-SA 2.0 (<https://www.flickr.com/photos/berniedup/53773387785>).

From a molecular perspective, the investigation of the structure and differentiation of these subspecies has been conducted solely at the level of a small region of the mitochondrial genome (Sithaldeen, Bishop, and Ackermann 2009). *P. ursinus* mitochondrial lineages cluster in two branches, mirroring geographic structure more than subspecies classification (Sithaldeen, Bishop, and Ackermann 2009). Typical chacma lineages are polyphyletic: one set of mitochondrial haplotypes from Southern and central South Africa group together, nestling a distinct clade comprising Ruacana lineages, while other Cape chacma mtDNAs, from the Northern region of South Africa, cluster with lineages of *Griseipes chacmas* (Sithaldeen, Bishop, and Ackermann 2009). According to the mitochondrial results, the demarcation line separating *Griseipes* and Cape chacma baboons appears located further down than the subspecies demarcation line, the North-Eastern Cape chacma clustering with *Griseipes*, while the Ruacana emerges as a subgroup nested within the variation of the South-Western Cape chacma (FIGURE 1.2B; Sithaldeen, Bishop, and Ackermann 2009).

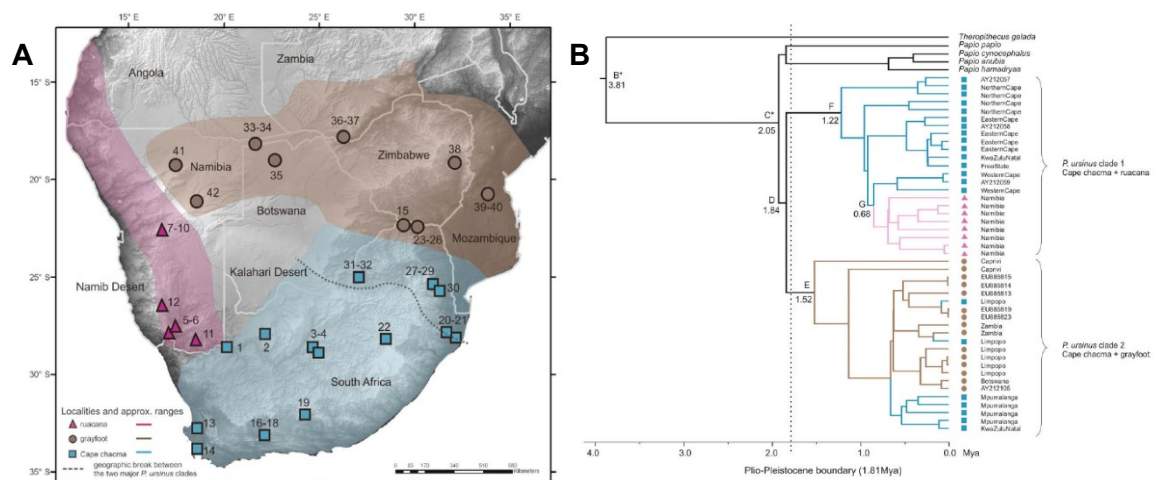


FIGURE 1.2 - *Papio ursinus* subspecies: A) distribution of the three subspecies (numbers and shapes refer to specimens in the original paper); B) mitochondrial phylogeny for the three subspecies according to the dataset in the original paper. Modified from Sithaldeen, Bishop and Ackermann 2009.

Analyses of the complete mitogenome are limited to two *P. ursinus* individuals, one from Southern South Africa (Kasanka National Park) and one from Northern South Africa (Blyde River; Zinner et al. 2013; Roos et al. 2021). In the context of *Papio* phylogeny, these two mitogenomes are in different parts of the phylogenetic tree, thereby confirming the polyphyly of the species. The Northern mitogenome branches alongside kinda and Southern yellow baboons, while the Southern mitogenome generates a branch external to the entire *Papio* phylogeny and is dated to 2.45Mya (Roos et al. 2021). A complete mitogenome, originating from an individual dating back 5,800 years, found in North-Eastern Lesotho, falls within the same branch as the Southern chacma, confirming a certain territorial and temporal continuity for this cluster (Mathieson et al. 2020). In very recent years, some genomes from the Northern portion of the chacma range have been released, and their haplotypes fall within the Northern chacma haplogroup, included in the clade comprising also yellow and kinda baboons (Santander et al. 2022; Sørensen et al. 2023).

Data on the male-specific Y chromosome system is very sparse for *P. ursinus*, so far only Y chromosomes belonging to individuals sampled in the Northern range of this species have been

analysed (Santander et al. 2022; Sørensen et al. 2023). As such, it is currently not possible to ascertain whether there is consistency with the mitochondrial data. However, both uniparental systems (Y chromosome and mitochondrial DNA) do agree in showing chacma analysed in Dendro Park in Zambia and Gorongosa National Park in Mozambique as sister taxon to yellow baboons (Santander et al. 2022; Sørensen et al. 2023; Kopp et al. 2023).

In Zambia and Mozambique, chacma baboons are known to interact with their neighbouring species, the kinda and yellow baboons, respectively (Jolly et al. 2011; Chiou et al. 2021; Martinez et al. 2019; Santander et al. 2022; Sørensen et al. 2023).

In Zambia, a hybrid zone has been identified in the region of the Kafue River National Park, located to the North of what is often referred to as the geographic boundary of the North-Eastern range of chacma baboons, the Zambezi River (Kingdon 1997). Individuals displaying hybrid characteristics between the two forms have been found in this area. Notably, individuals with chacma-like traits as well as individuals with chacma-kinda hybrids, possess a Y chromosome of kinda origin, although mitochondrial DNA is often of chacma origin. The prevalence of kinda Y chromosome in the hybrid zone suggests that kinda males have a reproductive advantage over chacma males. According to Jolly (2011), this can be explained by the fact that kinda males act as "sneakers", as they are smaller in size and are perceived by chacma males as females or, at most, as harmless juveniles. As a result, kinda males are able to approach chacma females, and through grooming (as the chacma females likely initially mistake them for other females), they form relationships that eventually lead to mating. Conversely, the occurrence of chacma males mating with kinda females appears to be infrequent, likely due to the significant size disparity between the two, and because chacma males probably perceive kinda females as too young and, therefore, unappealing (Jolly et al. 2011).

Beyond the Zambezi River, contact between chacma and yellow baboons has been identified in the Lower Luangwa Valley in Zambia by Burrell (2009). This interaction certainly extends eastward into central Mozambique, particularly in the area of Gorongosa National Park (Martinez et al. 2019). This finding is rather interesting, given that the Lower Zambezi in Mozambique is a very wide river, which baboons might be unable to swim across.

Martinez and colleagues (2019), in studying the morphological characteristics of Gorongosa baboons, observed a mosaic of traits shared with both Northern chacma and Southern yellow baboons. This suggests the presence of a North-South geographical cline, likely due to ancient or ongoing hybridisation. Given the baboons' inability to cross this geographical barrier, Martinez proposed that the Lower Zambezi River system was different in the past, allowing for secondary contact between yellow and chacma baboons, with Gorongosa becoming a hybridisation zone. Santander's work in 2022 detected genome signals of ancient hybridisation between chacma and yellow baboons, reinforcing the evidence of connections between the two species.

At the time of our publication of the first genomic data from Mozambican baboons, *Papio* whole genomes were restricted to a handful of individuals from the wild (Wall et al. 2016; Robinson et al. 2019; Rogers et al. 2019; Santander et al. 2022), preventing a more systematic search for the *P. cynocephalus* source population. Limitations also extended to the genomic variation sampled within *P. ursinus*. In fact, while a representative sampling of *Papio* mitogenome diversity highlighted the polyphyletic status of *P. ursinus* mitochondrial genomes (Zinner

et al. 2013; Roos et al. 2021), *P. ursinus* genomic data was limited only to two Gorongosa genomes (one with very low coverage; Santander et al. 2022) and two captive individuals of unknown provenance (Rogers et al. 2019). These results, while highlighting the peculiarity of baboons at the North-East boundary of *P. ursinus* range, left untested their affinity to different groups of *P. cynocephalus* and to what extent introgression was present in other groups of *P. ursinus*.

Recently, an extensive survey of *Papio* genomic variation comprising more than 200 individuals has been published (Sørensen et al. 2023). This investigation highlighted multiple hybridisation events between *Papio* species. In particular, the yellow baboons from Western Tanzania were described as the result of admixture between yellow, kinda and olive baboons (*P. anubis*), while the yellow baboons from Eastern Tanzania showed, in their genomes, only a little contribution from olive baboons. Unfortunately, despite investigating hundreds of *Papio* genomes, only four *P. ursinus* samples were included in this work. Interestingly, these samples originated from Dendro Park, in Zambia, approximately 1000 km West of Gorongosa and similarly at the Northern boundary of *P. ursinus* range extension. The analysis of Dendro Park *P. ursinus* revealed that this population is characterised by recent admixture with kinda and yellow baboons. The larger contribution of *P. cynocephalus* present on the X chromosome than the rest of the genome was interpreted as a male-biased admixture event. By now, it is not clear what of yellow baboon populations might be the sources of introgression in Central Mozambique and Zambia chacma.

Due to the absence of complete genomes of *P. ursinus* from various regions within its distribution, it has not yet been possible to conduct a comprehensive study of the population structure or assess the extent to which introgressions from other species have influenced the genome of this species. Currently, the available literature only includes genomes from hybridisation zones or zoo specimens, while individuals from the Southern or Western parts of the range remain unsampled.

In this chapter, the research aims to investigate the hybridisations between neighbouring species of baboons in Southern Africa and their influence on the population structure of *P. ursinus*. To achieve this goal, first of all, the investigation of the interaction between chacma from Central Mozambique and yellow was extended. Then the Zambia hybridisation zone was studied more in deep. Finally, the population structure of chacma baboons was investigated. Additionally, differential hybridisation signals were detected by characterising the directions and intensities of gene flow among the different chacma populations and the two neighbour species, *P. cynocephalus* and *P. kindae*.

1.2 - Materials and Methods

1.2.1 - Samples and Datasets

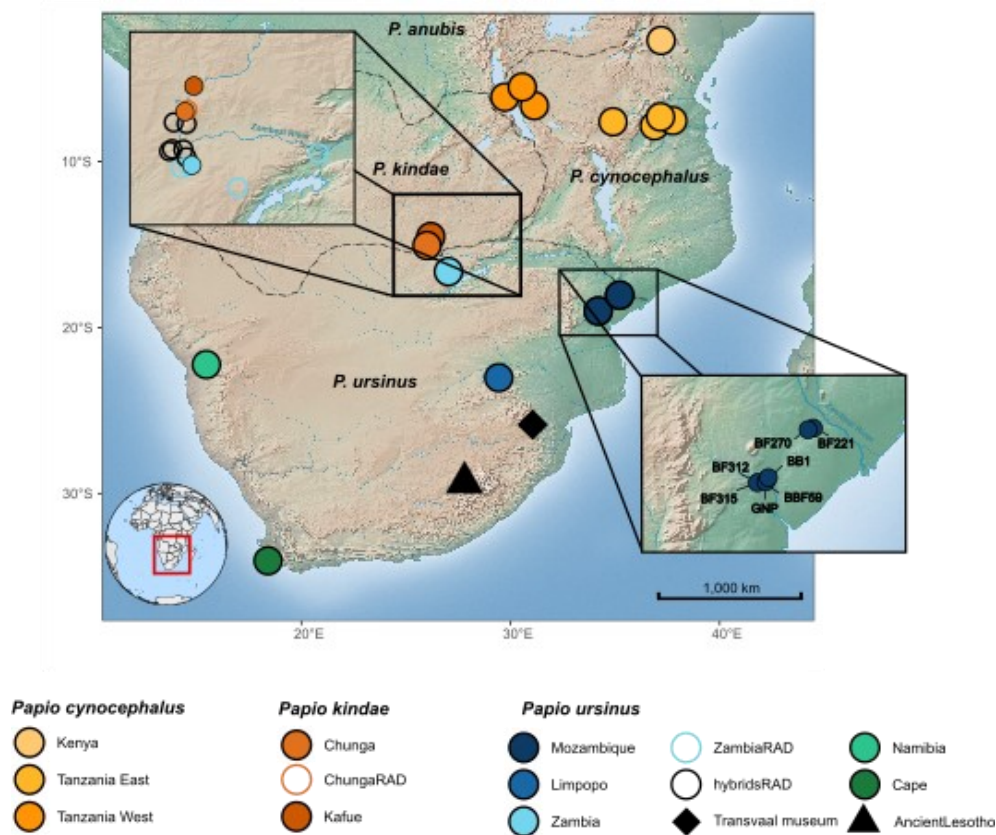


FIGURE 1.3 - Map with the provenience of sampled individuals. Colours indicating their respective modern populations. The black shapes represent the presumed origin sites for the ancient specimen (Mathieson et al. 2020) and for the museum specimen. The two insets provide an enlarged view of the origin locations for the samples from Zambia (Chiou et al. 2021; Sørensen et al. 2023) and central Mozambique. The dashed lines outline the species' range limits according to Sørensen et al. (2023). More details are in Supplementary Table 1.

In order to extend the available survey of *P. ursinus* variation, we sampled chacma baboons from locations in Southern Africa (Supplementary Table 1). This novel dataset included:

- Two adult baboons (one male and one female) from two different troops in Tsaobis Nature Park, Namibia (22°23'S;15°45'E; FIGURE 1.3), sampled in 2019 (Guy Cowlshaw, Dada Gottelli, Alecia Carter and Elise Huchard; (Martina, Cowlshaw, and Carter 2020; Mutti et al. 2023); DNA was shotgun sequenced at Wellcome Trust Centre for Human Genetics (Oxford, UK) with a final coverage of 37.83X and 35.11X. From this point onwards, both samples will be considered as a single population, which will be referred to as “Namibia” (Supplementary Table 1).
- Two adult males from the Cape region (FIGURE 1.3), one from the Tokai Plantation (Table Mountain National Park, South Africa - 34°05'S;18°40'E) and one from Plateau Road which runs adjacent to the Cape of Good Hope Reserve, South Africa (34°23'S;18°40'E), collected in 2020 (Vincent Naude). From this point onwards, both

samples will be considered as a single population, which will be referred to as “Cape” (Supplementary Table 1).

- Two adult males sampled in the Western Soutpansberg Mountains in the Limpopo region, South Africa (29°44'S;23°02'E; FIGURE 1.3) from the same troop in 2018 (Russel Hill and Cyrintha Joubert; (Allan, Bailey, and Hill 2020). From this point onwards, both samples will be considered as a single population, which will be referred to as “Limpopo” (Supplementary Table 1).

The Cape and Limpopo samples were sequenced using PacBio sequencing (HiFi) by Inqaba Biotech to a coverage respectively of 16.95X and 18.58X, 40.49X and 24.46X.

We also analysed the genome data obtained from a specimen conserved at the Natural History Museum Giacomo Doria (Genova, Italy). The specimen is represented by a skull (MSNG32161; FIGURE 1.3-1.4) with a very damaged museum label, revealing “*Papio porcarius*, Boddaert. Transvaal: Swaziland”. *Papio porcarius*, Boddaert was one of the early names for *Papio ursinus*. This specimen was collected in 1931 and comes from the Eastern portion of the former Transvaal region, now the province of Mpumalanga, or the neighbouring Kingdom of Eswatini. For simplicity, it will be referred to by the locality recorded in the museum archives as “Transvaal” (Supplementary Table 1). DNA from the museum specimen was recovered and sequenced as described in (Saupe et al. 2021) by the Institute of Genomics, University of Tartu, Estonia to a final coverage of 0.72X.

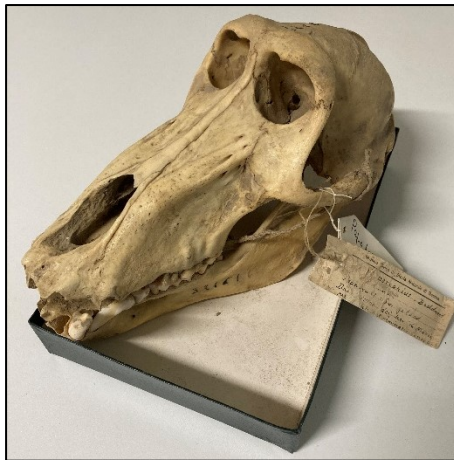


FIGURE 1.4 - *Papio ursinus* skull (male; MSNG32161). From Natural History Museum “Giacomo Doria” in Genoa. Photo by Matteo Caldon

We finally selected a subset of 5 faecal samples over the 400 collected in 2017-2018 by Maria J. Ferreira da Silva in Gorongosa National Park and Catapú Reserve Forest (150 km North of the Park; FIGURE 1.3). Samples selected were sequenced on an Illumina platform, reaching the mean coverage across samples of 0.0066X (Supplementary Table 1).

The chacma genomic dataset was completed by including all the available chacma genomes (FIGURE 1.3; Supplementary Table 1): two from Central Mozambique (Santander et al. 2022); four from Dendro Park in Zambia (which will form the population referred to as “Zambia”;

Sørensen et al. 2023); the genomic data recovered from a 5800-year-old chacma specimen from Lesotho (which will be referred to as "ancient Lesotho specimens" or simply "ancient"; Mathieson et al. 2020).

Finally, 54 genomes generated with RAD sequencing from Zambian faeces were added (Chiou et al. 2021). Those individuals are from various localities of the hybridisation zone in Central Zambia, some are identified as chacma-kinda hybrids by Chiou and colleagues (2021).

The chacma dataset was further completed by adding a set of genomes of baboons from the Southern *Papio* clade (FIGURE 1.3; Supplementary Table 1), selected according to their original location. We started with the broadest genomic dataset available for *Papio* (Sørensen et al. 2023) and selected from each location two individuals, resulting in five and eight yellow baboons from, respectively, Western Tanzania (Issa Valley, Mahale and Katavi; which will form the population referred to as "Tanzania West") and Eastern Tanzania (Mikumi, Ruha, Udzungwa, Selous; which will form the population referred to as "Tanzania East"; Supplementary Table 1) and three kinda baboons from Chunga (Zambia; which will form the population referred to as "Chunga"). To this selection, we added a group of other individuals with high-coverage genomes from locations not sampled in the Sørensen dataset: eight yellow baboons from Southern Kenya (founders of the SNPRC colony; Robinson et al. 2019), one from Amboseli National Park, Kenya (Wall et al. 2016), which will form the population referred to as "Kenya"; three kinda baboons from the Kafue National Park in Zambia (Rogers et al. 2019) which will form the population referred to as "Kafue" and 69 RAD sequencing kinda from Chunga in Zambia (Chiou et al. 2021).

The genome dataset was finally integrated with an outgroup, *Mandrillus leucophaeus* (Drill) recovered from PRJNA785018 BioProject in NCBI.

1.2.4 - Genome Mapping

All the reads were mapped with BWA-MEM2 v.2.2.1 (Vasimuddin et al. 2019), to the reference assembly genome Panubis1 (Batra et al. 2020). We then marked the duplicate reads with Picard MarkDuplicates version 2.8.1 (<http://broadinstitute.github.io/picard/>) and filtered the result using the option -q 10, -F 1292 and -f 2 of samtools view (Danecek et al. 2021).

1.2.5 - Variant Calling and Final Dataset

Following GATK best practices, we call the autosomal chromosome variants, using the GATK version 4.2.4.1 (McKenna et al. 2010). For each sample, we generated the gVCF with HaplotypeCaller, and we made the joint calls via GenotypeGVCFs. The VCF thus obtained was filtered with VariantFiltration using hard filters for SNPs ("QD < 2.0 || MQ < 40.0 || FS > 60.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0").

With PLINK 1.9 (Chang et al. 2015), all sites from the sex chromosomes were removed, keeping only the biallelic SNPs filtered using the --snps-only option. The final set of analysed markers included 28,627,144 autosomal SNPs. The SNPs panel was used to call the variants with pileupCaller (Lamnidis et al. 2018; <https://github.com/stschiff/sequenceTools>) of ancient, museum and low-coverage samples.

1.2.6 - Autosome Analysis

PCA

We performed a preliminary survey of population structure in the set of samples of Southern *Papio* here analysed by running a principal component analysis (PCA) using the smartpca function implemented in EIGENSOFT software 8.0.0 (Patterson, Price, and Reich 2006). We performed these PCAs for the high-coverage samples (Supplementary Table 1). The low-coverage, RAD, ancient and museum specimens were instead projected using lsqproject set to YES. All the other settings were used with default parameters.

ADMIXTURE

To understand the genomic components, present in individuals sampled in different locations of Southern Africa we performed ADMIXTURE analysis using the software ADMIXTURE v.1.3.0 (Alexander, Novembre, and Lange 2009), applying the --cv option to estimate the cross-validation values, exploring K values from 2 to 6.

Heterozygosity

To calculate the proportion of heterozygous sites (HET) we used PLINK --het. This command estimated the observed O(HOM) homozygous genotype counts for each sample and the total number of variable sites (N_Sites), from which we derived HET per individual with the formula:

$$\text{HET} = \frac{N_{\text{sites}} - O(\text{HOM})}{N_{\text{Sites}}}$$

D-statistics

We performed the D-statistics (ABBA-BABA test) using qpDstat (Patterson et al. 2012), with default parameters, to test for the occurrence of any imbalance in allele sharing between different populations and species (H1, H2 and H3), including drill (*M. leucophaeus*) as an outgroup (H4). The test considers the ancestral alleles (A) and the derived alleles (B) in the genomes of the four taxa considered. Without gene flow, the two conflicting allelic patterns (ABBA and BABA) are equally frequent, and the values of D are zero. When there is gene flow, there is an excess of one of the two allelic patterns which produces a deviation of the D from zero: positive D (excess of ABBA) indicates the introgression between H2 and H3, and negative D (excess of BABA) points to the gene flow between H1 and H3. The observations are considered significant if $|Z|$ -score > 3.

f4-ratio

To infer the mixing proportions of an admixture event between the chacma baboon populations and the other two baboon neighbour species, we calculated the f4-ratio through qpF4ratio implemented in AdmixTools (Patterson et al. 2012) including drill (*M. leucophaeus*) as an outgroup (O).

PSMC

Demographic changes (effective population size history) over time were explored using the Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011). As

recommended, we generated diploid sequences using bcftools (Li 2011); sites with coverage greater than double the average genome-wide coverage or less than one-third were excluded. We also excluded the repetitive regions in the genomes, as indicated on the NCBI website (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_003255815.1).

We run PSMC with default settings, using a mutation rate of 0.9×10^{-8} per base pair per generation and a generation time of 11 years (Santander et al. 2022). Finally, we performed bootstrapping for each individual with 10 replicates following the instructions described on the GitHub webpage manual (<https://github.com/lh3/psmc>).

In a similar manner, this software can also be used to explore the last time of contact between populations by creating artificial F1 hybrids (hPSMC; Cahill et al. 2016). Following the procedure described on the GitHub webpage manual (<https://github.com/jacahill/hPSMC>), we generated from the bam a haploid sequence in fasta format of the X chromosome for each male individual. We then combined the fasta sequences from two different individuals into a single .psmcfa file, and then we applied PSMC with default settings. We use a mutation rate of 0.419×10^{-8} mutations per base pair per generation. It became necessary to recalculate the autosomal mutation rate, as the X chromosome, due to its presence in different numbers in male and female individuals, results in a different mutation rate. Therefore, the autosomal mutation rate was converted into X mutation rate as described by Li and Durbin (2011), using the male-to-female mutation ratio of *Rhesus macaque* (2.9; Wilson Sayres et al. 2011).

1.3 - Results

1.3.1 - Preliminary Analysis of Chacma Baboons at the Northern Border of the *P. ursinus* Range: Zambia and Mozambique

Chacma baboons in Gorongosa National Park, Mozambique

The population structure and gene flow between the populations of *P. ursinus* at the boundary with kinda and yellow baboons were initially analysed including data from the Sørensen (2023) dataset, along with the inclusion of newly generated low-coverage samples from central Mozambique and the previously published high-coverage sample from the same area (Santander et al., 2021).

The PCA of autosomal genome-wide markers highlighted three main clusters, mostly reflecting the three species (FIGURE 1.5A). The main exception was *P. cynocephalus* baboons from Western Tanzania (Issa Valley, Mahale, Katavi), which appeared closer to *P. kindae* than *P. cynocephalus* from Eastern Tanzania (Mikumi, Ruha, Udzungwa, Selous), as previously reported (Sørensen et al. 2023). *P. ursinus* populations from Mozambique and Zambia clustered close to each other. The low-coverage genomic samples recovered from faecal material formed a cloud around the high-coverage bf186 and the low-coverage BB1 samples previously published, suggesting a general genomic similarity for samples from Gorongosa National Park and Catapù Reserve Forest (FIGURE 1.5A).

We then tested, through D-statistics, whether individuals with low coverage and individual with high coverage from Mozambique were equally related to other species and to the chacma population in Zambia. We observed that overall, with some degree of interindividual variation, samples tend to behave similarly in relation to other species and populations (FIGURE 1.5B). For this reason, we decided to consider all individuals from Gorongosa National Park and Catapù Reserve Forest as a single population in subsequent analyses, which will be referred to as ‘Mozambique’ for simplicity.

The patterns of allele sharing of the two populations of *P. ursinus* with the two groups of *P. cynocephalus* highlighted a broad similarity between the chacma groups (FIGURE 1.5C). Both share more alleles with Eastern than Western yellow baboons ($D=0.0063$ $Z=3.293$, for Zambia; $D=0.0291$ $Z=14.257$, for Mozambique). However, when we compared the two populations with *P. kindae* the pattern was different, the Zambian population shared more alleles with this species than the Mozambican one ($D=0.0663$ $Z=32.130$). Similarly, *P. ursinus* in Zambia was closer to Western yellow baboons than *P. ursinus* in Mozambique ($D=0.0300$ $Z=19.645$; FIGURE 1.5D).

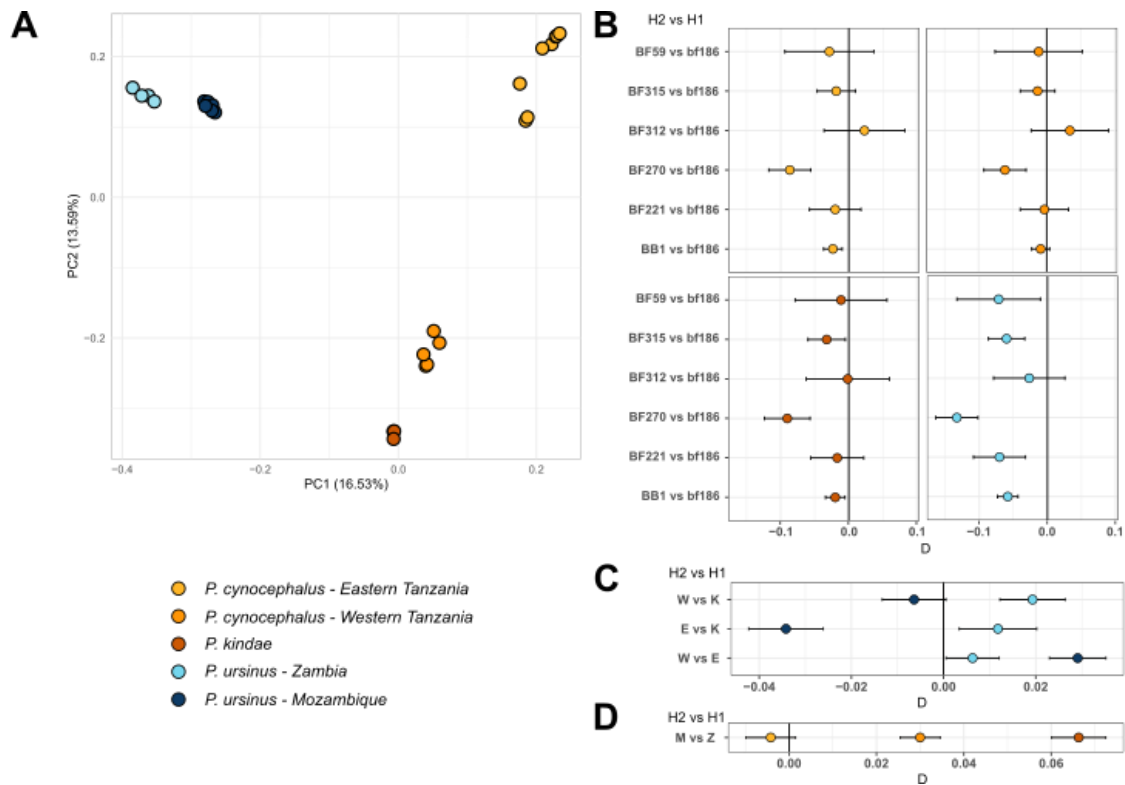


FIGURE 1.5 - Chacma baboons in Gorongosa NP: A) PCA of autosomal genome-wide markers; B) D-statistics comparing the high coverage (bf186 as H1) and the low coverage (BF59, BF221, BF270, BF312, BF315, BB1 as H2) individuals from Gorongosa National Park to neighbouring taxa (*P. cynocephalus* from eastern Tanzania, *P. cynocephalus* from western Tanzania, *P. kindae* and *P. ursinus* from Zambia; H3); C) D-statistics comparing neighbouring species (*P. cynocephalus* from eastern Tanzania, *P. cynocephalus* from western Tanzania and *P. kindae* as H2 and H1) to chacma populations (Mozambique and Zambia as H3). D) D-statistics comparing chacma populations (same as C; as H2 and H1) to neighbouring species (same as B; as H3). In B, C and D, the bars show the extension of three standard deviations and the colours refer to the taxon used as H3 as indicated in the legend; In C and D, the H1 and H2 labels refer to *P. cynocephalus* from eastern (E) Tanzania, *P. cynocephalus* from western Tanzania (W), *P. kindae* (K), *P. ursinus* from Zambia (Z) and *P. ursinus* from Mozambique (M). For positive values, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

Chacma baboons in Zambia

In a subsequent phase, chacma and kinda from Zambia were considered with samples obtained from faecal material through RAD sequencing (Chiou et al. 2021) and the kinda from Kafue National Park (Rogers et al. 2019). This section thus explores the structure and relationships between the populations of the two forms from the Zambia region, where hybridisation events have been reported.

The PCA of autosomal SNPs suggests a separation into two primary groups along principal component 1 (34.66%), which identifies the two species in analysis, *P. ursinus* and *P. kindae* (FIGURE 1.6A). At one end, all samples from Dendro Park (four high-coverage samples and three of four RAD samples) are found, while at the opposite end, all kinda samples are observed. Between these two groups, there is a gradient of individuals previously identified as chacma or hybrids, creating a continuum between the two forms. Principal component 2 (9.53%), shows separation along kinda samples. Most are located in a central cluster, while a few are positioned above and below this cloud. The kinda samples from Rogers (2019) and Sørensen (2023) do not cluster uniquely; some fall within the cluster, while others are positioned above or below the main cloud.

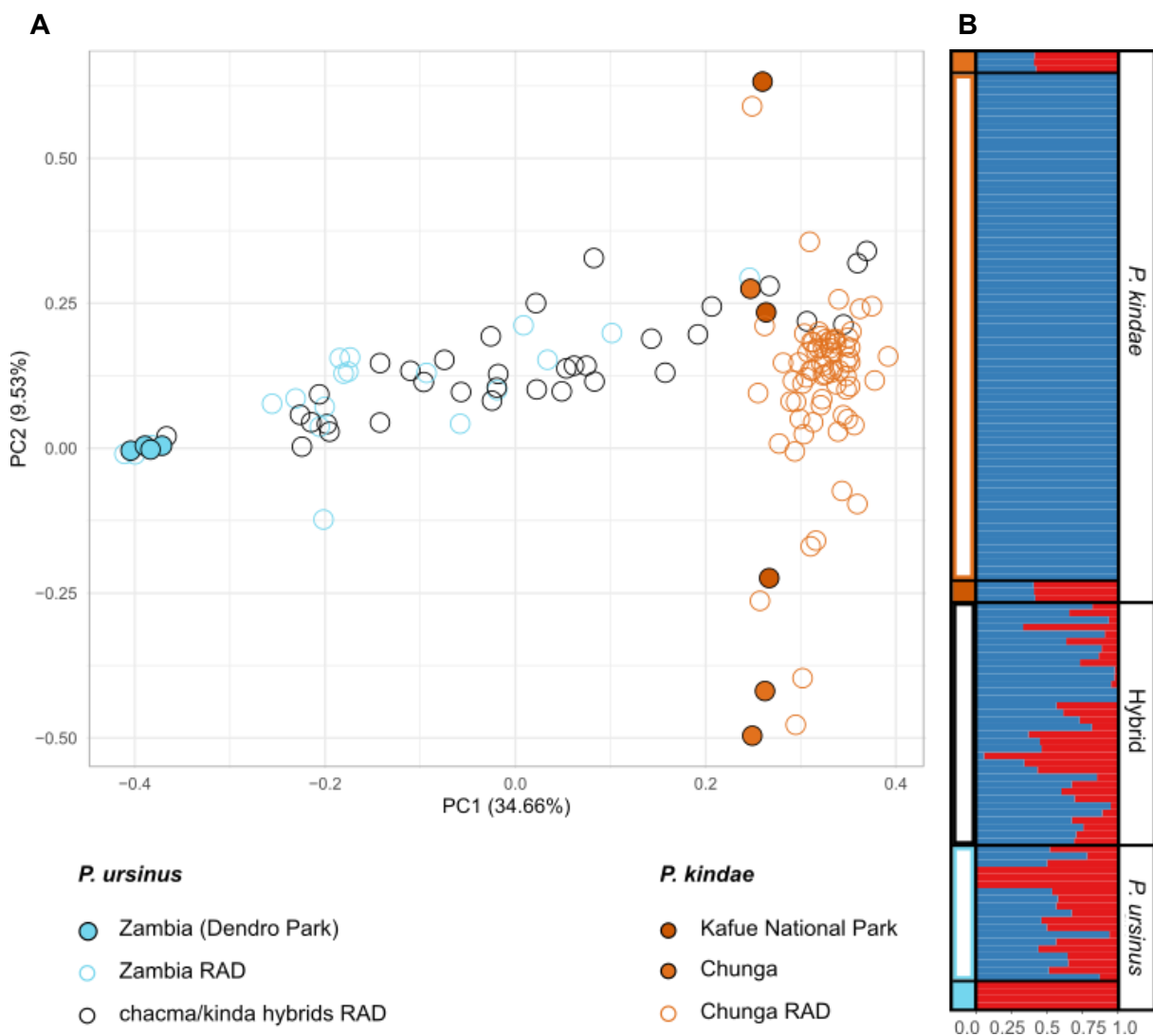


FIGURE 1.6 - Population structure of baboons in Zambia A) PCA of autosomal genome-wide markers; B) ADMIXTURE analysis for K=2 (CV=0.43151).

ADMIXTURE analysis indicated that the most supported number of K values was K=2 (CV=0.43151; FIGURE 1.6B), mirroring the two species here analysed *P. ursinus* and *P. kindae*. The high-coverage chacma and other RAD samples from Dendro Park are composed of a single component (red in FIGURE 1.6B), while all kinda RAD samples consist entirely of the other component (blue in FIGURE 1.6B). All remaining individuals exhibit varying proportions of both components, including the high-coverage kinda from Chunga and Kafue. As observed in the PCA, many individuals assigned to chacma show a genomic composition comparable to that of individuals classified as hybrids.

Using the f4 ratio, we tested the proportion of chacma (and kinda) components in each of the chacma and hybrid RAD individuals (FIGURE 1.7). This analysis quantifies the degree of affinity of population X with population B relative to population C. Cape population was used as A; as C, the two kinda populations (Chunga and Kafue); as B, Zambia, Limpopo, and Namibia; and as X, all chacma individuals from Zambia (including hybrids, those assigned to chacma, and with only B as Limpopo and Namibia, also the Zambian population from Sørensen's dataset). When individuals were tested using Zambia as B, it was observed that the contribution of the Zambian population to these individuals varied greatly, from just above zero to 100%. In some cases, the values even exceeded 1, indicating that the tree structure is likely incorrect, as the tested individuals had probably less contribution from kinda than the Zambian population in B.

When other available chacma populations from Limpopo and Namibia were alternatively tested as B, it was observed that the chacma contribution values in each individual were lower than previously observed, suggesting that Namibia and Limpopo likely have less kinda than Zambia. In fact, when Zambia is included as the test population (X), the result of the analysis shows about 60% chacma contribution (FIGURE 1.7). With Namibia or Limpopo as population B, the individuals who previously had unsupported values ($\alpha > 1$) now have supported values ($0 < \alpha < 1$),

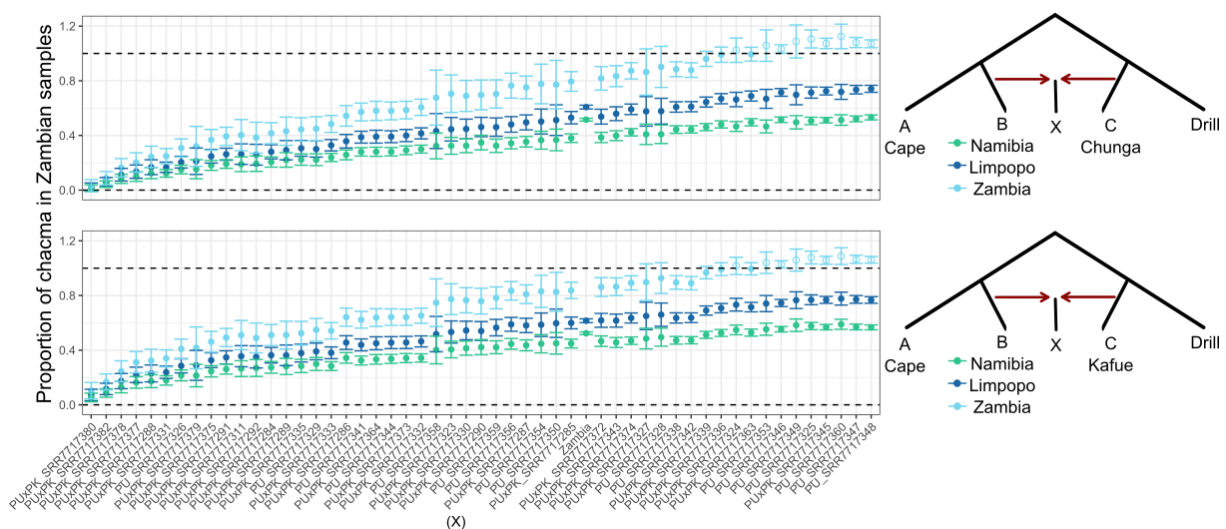


FIGURE 1.7 - Proportion of chacma component in Zambia baboons: f4-ratio tests for chacma and chacma/kinda hybrids. The tree topology used is the one indicated on the right. The colours refer to the chacma population used as branch B. The empty dots refer to alpha values associated to unsupported Z values; the bars show the extension of three standard deviations.

but still mostly higher or equal to those of the Zambian population. These results are entirely similar whether Kafue or Chunga kinda is used as C.

Due to the variation in the chacma component, it was decided not to include these RAD samples in the Zambia population. As for the kinda, the Kafue and Chunga groups will be kept as separate populations.

1.3.2 - Population Structure of Chacma Baboons

After observing a differential gene flow between the chacma baboons in the Northern part of the species range and nearby species, and detecting some structure between the two populations considered, the dataset of chacma baboons was expanded. This included the newly generated samples from Namibia and from two different locations in South Africa, (from the Southern region of the Cape of Good Hope and from the Northern region of Limpopo). Additionally, an ancient sample from Lesotho, dated 5,800 years ago (Mathieson et al. 2020), and a newly generated museum sample from the Mbupalanga region/Eswatini kingdom, conserved at the Genoa Museum, were incorporated. The dataset for other species was also broadened by including a new yellow baboon population from Kenya composed of the founders of the SNPRC colony (captured in Southern Kenya; Robinson et al. 2019) and a sample from Amboseli National Park in Southern Kenya (Wall et al. 2016; see Materials and Methods and Supplementary Table 1).

The degree of population structure present across Southern *Papio* species was initially explored using Principal Component Analysis. The chacma individuals are separated along the first principal component (18.69%) from yellow and kinda baboons, the latter separated from each other along the second principal component (11.89%; FIGURE 1.8A). Along PC2 the yellow baboons from Western Tanzania are positioned between yellow from Eastern Tanzania/Kenya and the Kinda baboons from Zambia. Chacma from Cape are located in one corner of the PCA plot, the other *P. ursinus* samples moving towards the centre of the plot, first individuals from Namibia and Limpopo, and then those from Zambia and Mozambique. Thus, four distinct groups are evident among chacma, one comprising individuals from the South and West (Cape and Namibia), the other two consisting of individuals from the North-central and North-East respectively (Zambia and Mozambique) and one with individuals from the Limpopo, positioned in between. The ancient individual from Lesotho (Mathieson et al. 2020) aligns with the South and West chacma groups while the historical museum specimen aligns with individuals from Limpopo, indicating a certain geographic coherence. Low-coverage baboon samples from Mozambique form a cloud around the high-coverage Mozambican sample, while the hybrids and chacma samples from Zambia generated a continuum with kinda, as reported above (FIGURE 1.5A - 1.6A). In this analysis, the kinda samples group more closely together than reported in FIGURE 1.6A.

The running of ADMIXTURE using genomic data of Southern baboons indicated $K=3$ ($CV=0.22686$) as the best fitting model, the three suggested clusters aligning with each one of the three Southern *Papio* species *P. ursinus*, *P. cynocephalus* and *P. kindae* (FIGURE 1.8B). The only exceptions were chacma samples from Mozambique and Zambia that showed association with more than one ancestry. Interestingly these two populations differ from each other in their ancestry assignment: Zambian baboon profiles belong prevalently to the chacma ancestry with

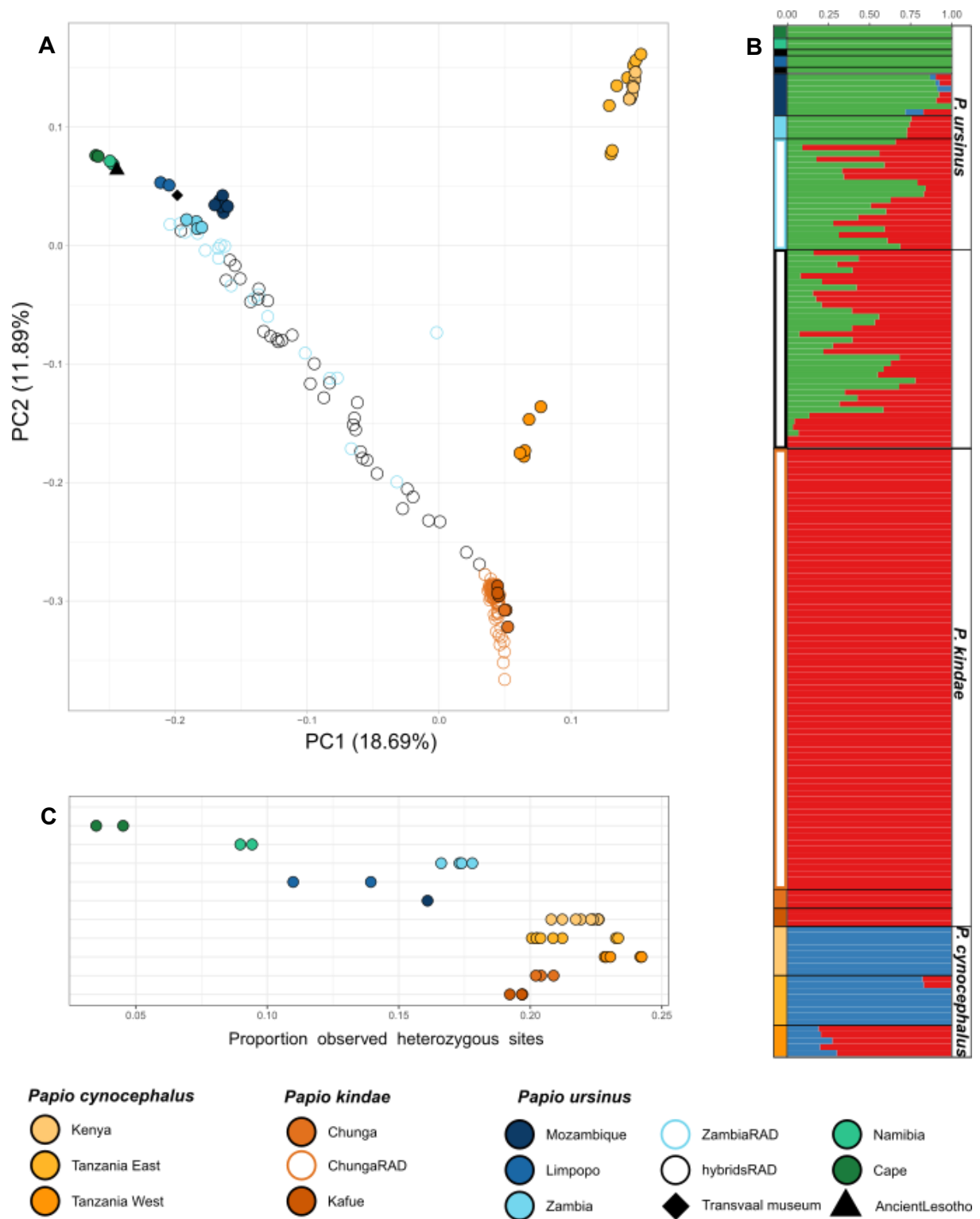


FIGURE 1.8 - Population structure of chacma baboons. A) PCA of autosomal genome-wide markers; B) ADMIXTURE analysis with $K=3$ ($CV=0.22686$); C) estimation of the proportion of observed heterozygous sites. In A, B and C the colours refer to the population colour in the legend

a portion of kinda ancestry, while the Mozambican baboon profiles are composed predominantly by the chacma, with minor ancestry affiliations to both kinda and yellow. As previously reported the Zambian samples presented a variable proportion of chacma and kinda in their genomic composition. Unlike the ADMIXTURE involving only Zambia individuals, kinda baboons from Sørensen (2023) and Rogers (2019) datasets exhibit 100% kinda component, just like all the other kinda baboons considered. It is also observed that all the yellow baboons from Western Tanzania display a large percentage of kinda ancestry, while only some from the East show a kinda component, though much more limited.

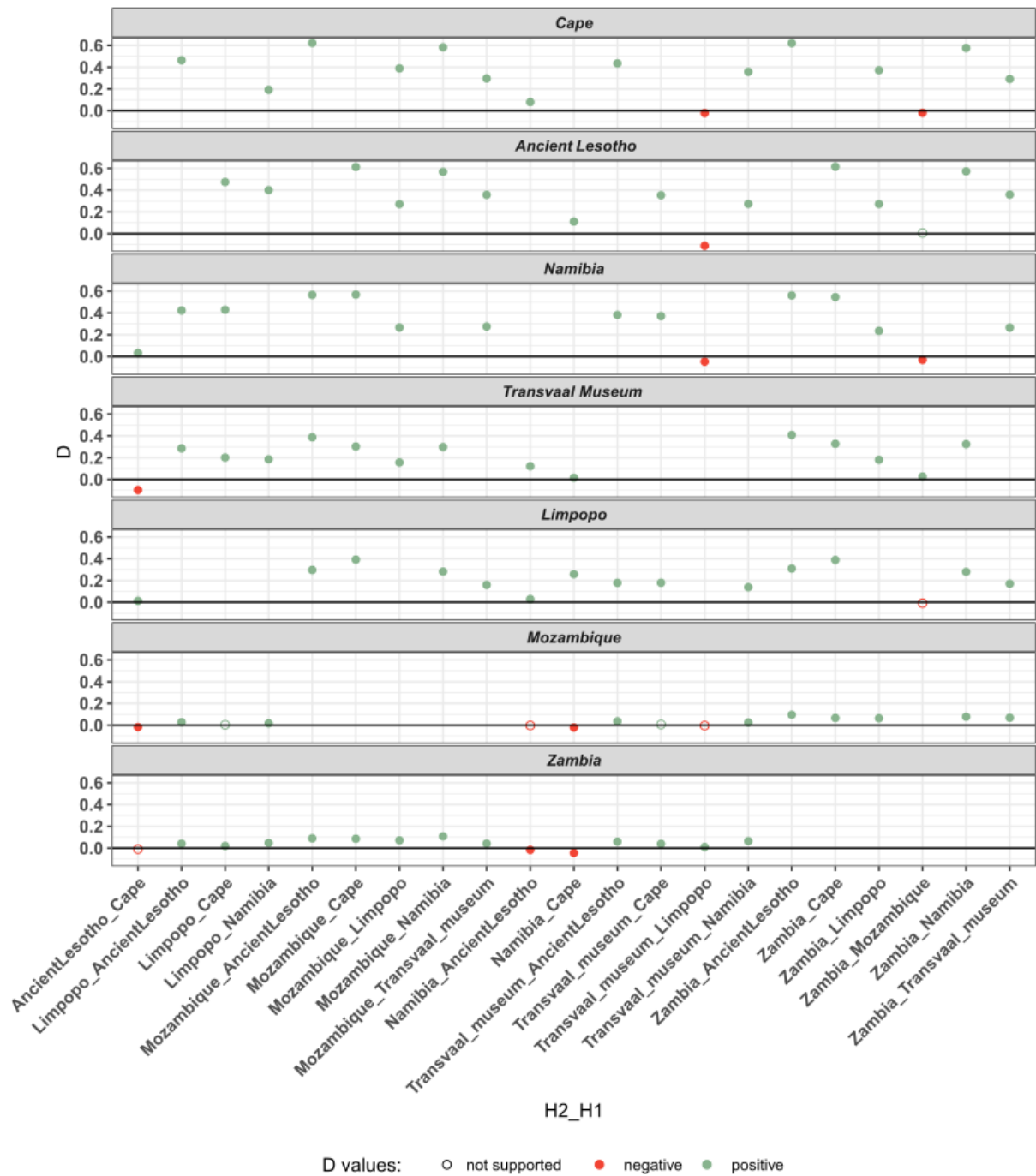


FIGURE 1.9 - D-statistics comparing chacma baboons. The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. The population name at the top of each panel refers to the chacma population used as H3. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

Then, the proportion of heterozygous positions in Southern baboons was estimated (FIGURE 1.8C). The Cape chacma is the population with the lowest heterozygosity, followed by those from Namibia and Limpopo. Populations from Mozambique and Zambia are the chacma with the highest proportion of heterozygous sites. Kinda and yellow baboons have a higher proportion of heterozygous positions than *P. ursinus*; the population with the higher heterozygosity is the one of yellow baboons from Western Tanzania. Among kinda, the Kafue population had a lower proportion of heterozygosity than the Chunga one.

Finally, a formal evaluation of the relationships among chacma populations using D-statistics was then performed (FIGURE 1.9). This analysis shows that Namibia, Limpopo and Cape (as the ancient and the museum sample) exhibit similar patterns of allele sharing among pairs of populations but different from those displayed by Zambia and Mozambique, which show similar patterns to each other. Notably, Cape and Namibia share more alleles between them than they do with Limpopo. Overall, two groups can be detected, Zambia and Mozambique on one side and all the other *P. ursinus* populations (including ancient and museum specimens) on the other.

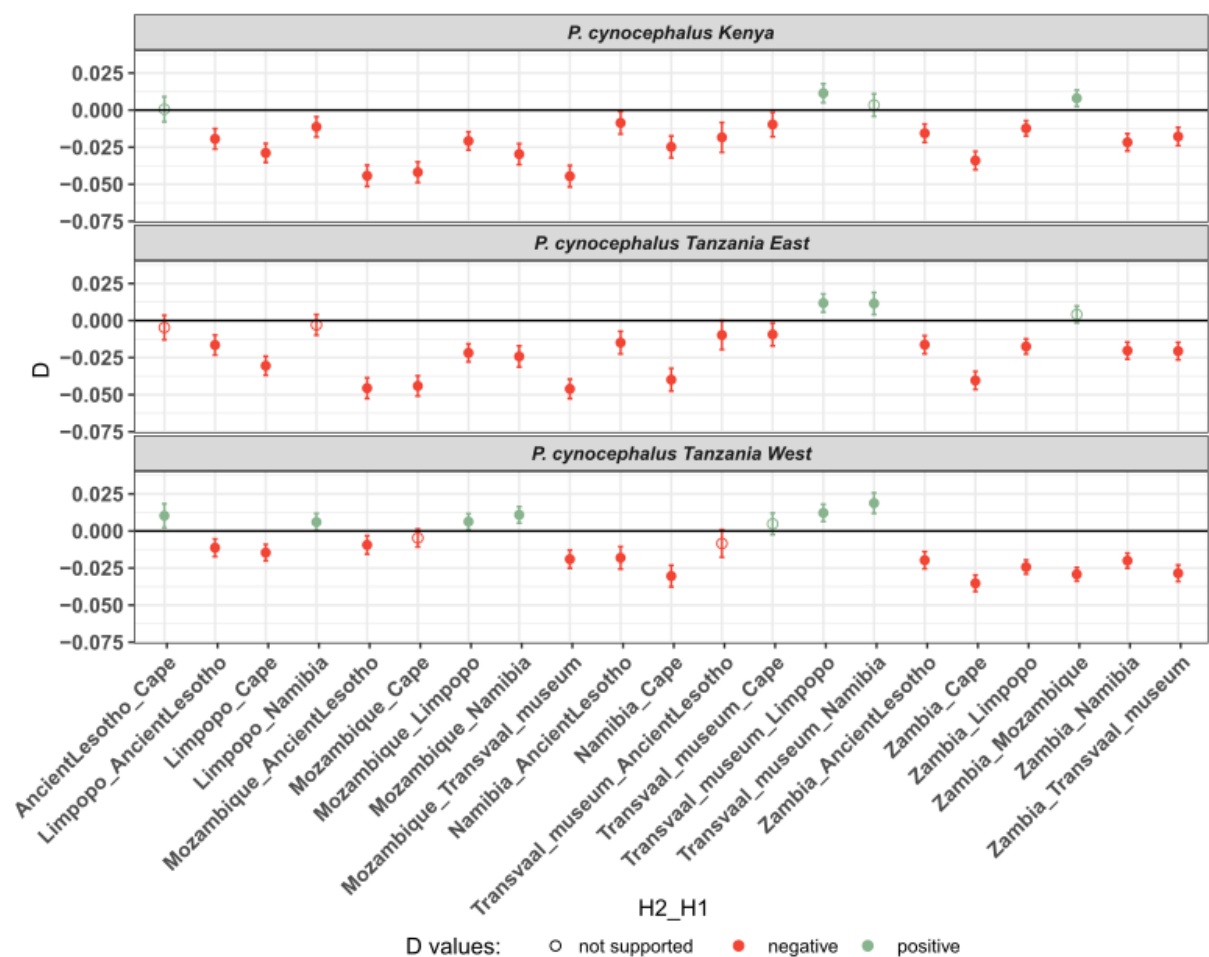


FIGURE 1.10 - D-statistics comparing chacma and yellow baboon populations. Chacma populations are used as H1 and H2, yellow populations as H3. The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. The population name at the top of each panel refers to the yellow population used as H3. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

1.3.3 - Relationship Between *P. ursinus* and Neighbouring Species

The occurrence of imbalance in allele sharing between chacma and yellow/kinda baboons was formally tested using the D-statistics. The difference in allele sharing between yellow populations (as H3) and chacma populations (as H2 and H1) was initially tested (FIGURE 1.10). The results suggest that the Cape and ancient Lesotho are the ones with the least allelic sharing with yellow baboons when compared to all the other chacma populations, while the Mozambique and Zambia populations have the strongest. The pattern is similar between yellow Kenya and yellow Tanzania East, with three exceptions: Limpopo/Namibia which is not significantly different for Tanzania East and on the side of Limpopo for Kenya; Namibia/museum which is not significantly different for Kenya, while is on the side of Namibia for Tanzania East; Zambia/Mozambique which is neutral for Tanzania East, and on the side of Mozambique for Kenya. The emerging general pattern for Tanzania East and Kenya is that the strongest allele sharing is with Mozambique (and Limpopo), followed by Zambia, Namibia and finally Cape. On the other hand, results suggest that Tanzania West has a stronger allele sharing with Zambia followed by Namibia, Limpopo and Mozambique/Cape.

The D-statistics (FIGURE 1.11) computed between kinda populations (as H3) and chacma populations (as H2 and H1) suggest that kinda from Kafue and kinda from Chunga are quite different (although they come from the same region). Kinda from Kafue, seems to affect more Mozambique, while kinda from Chunga seems to share more alleles with Zambia. The patterns

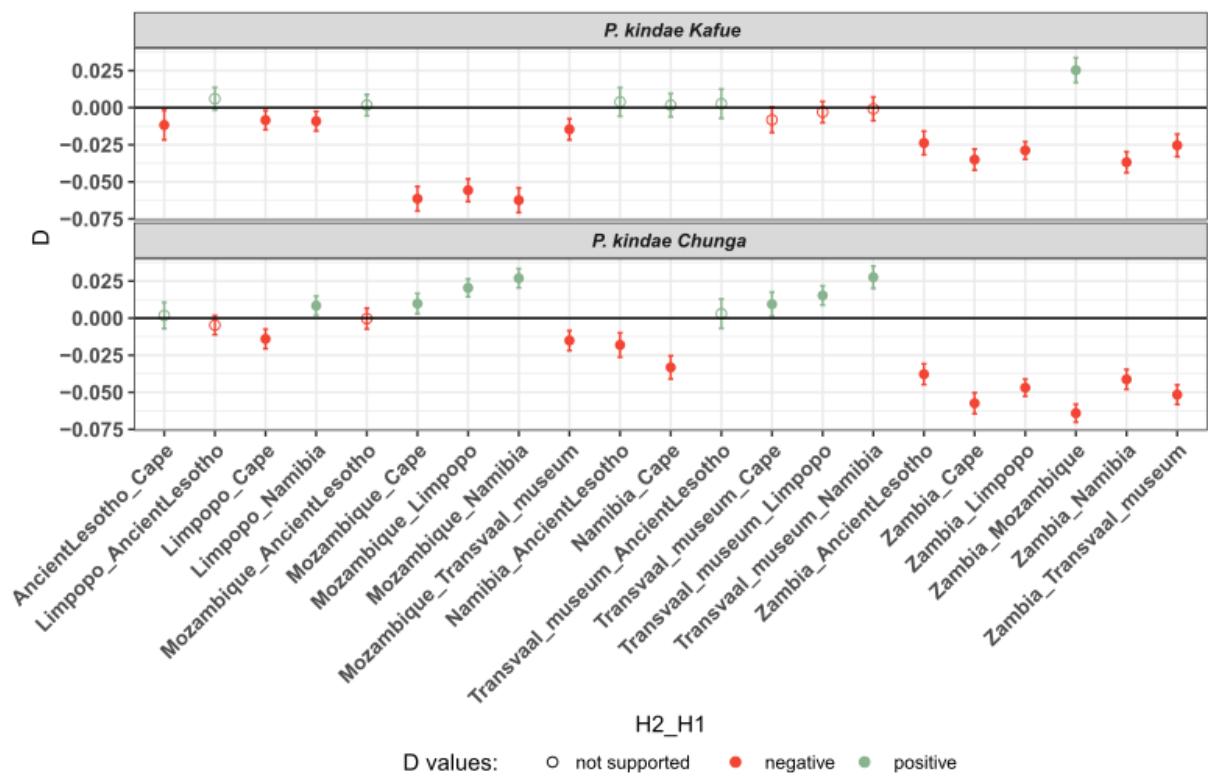


FIGURE 1.11 - D-statistics comparing chacma and kinda baboon populations. Chacma populations are used as H1 and H2, kinda populations as H3. The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. The population name at the top of each panel refers to the kinda population used as H3. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

for Limpopo and Namibia follow those of Mozambique and Zambia, respectively (Namibia to kinda from Chunga, Limpopo to kinda from Kafue). Lesotho and Cape are the chacma populations with lower allele sharing with kinda.

After this was tested D-statistics using chacma populations as H3 and populations of yellow and/or kinda as H2/H1 (FIGURE 1.12). The results suggest that all chacma populations have

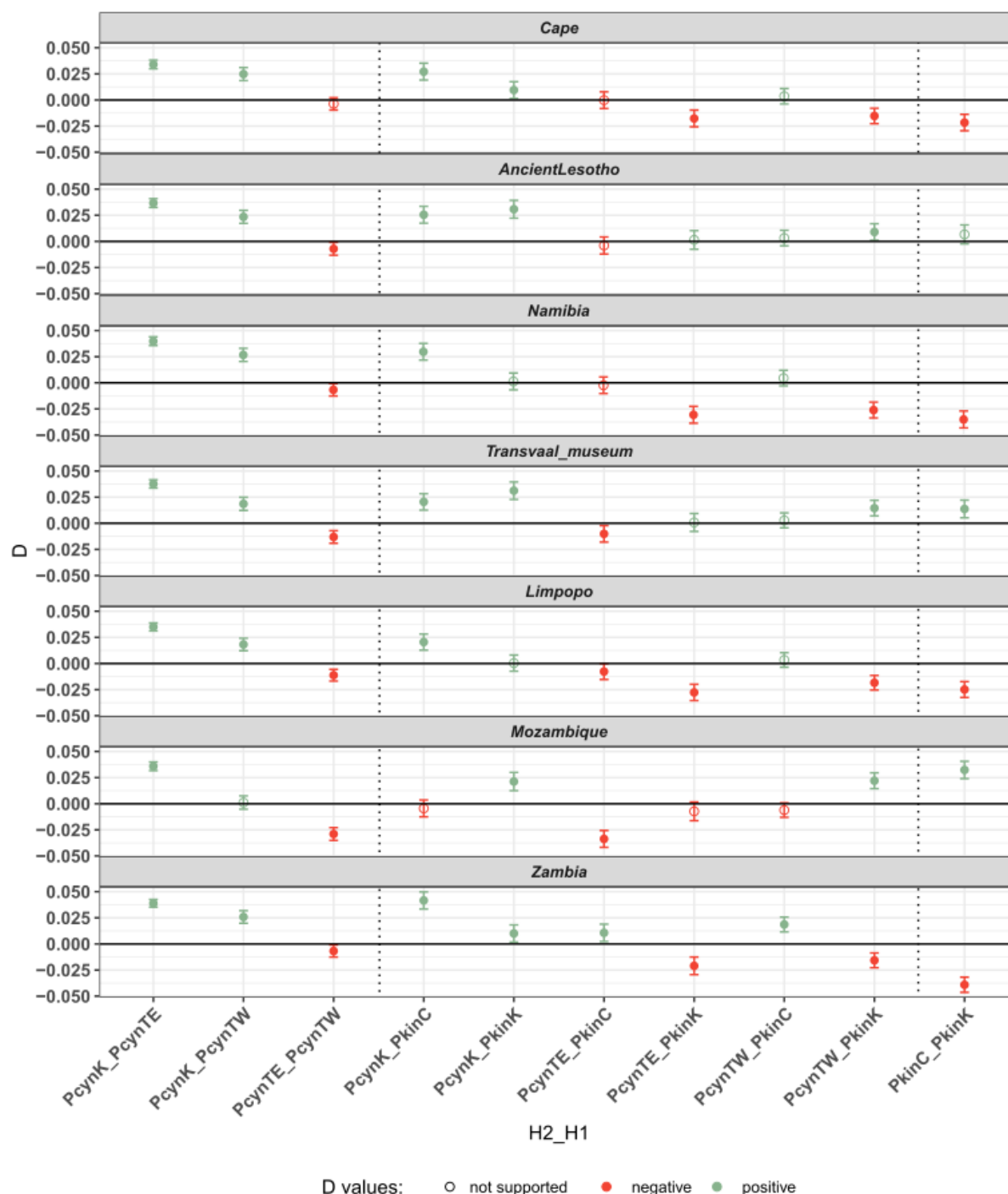


FIGURE 1.12 - D-statistics comparing yellow and kinda populations with chacma baboon populations. Chacma populations are used as H3, kinda and yellow populations as H1 and H2. The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. The population name at the top of each panel refers to the chacma population used as H3. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

less allele sharing with the Kenyan population than with the two yellow populations from Tanzania, but there are some differences. Tanzania East shows greater allele sharing with chacma populations, but the signal was stronger for Mozambique and for Limpopo, while for Zambia and Namibia was very near to zero (but still supported). The allele imbalance between Tanzania West and Kenya is always on the side of Tanzania West, with the only exception of Mozambique for which the comparison is neutral. When kinda from Kafue and kinda from Chunga were tested as H2 and H1, the results suggest that kinda Chunga share more alleles with chacma populations than the other, with the exception of Mozambique and museum specimens, while for the ancient specimens from Lesotho, the comparison is neutral (FIGURE 1.12).

Both kinda populations share more alleles with chacma populations than yellow Kenya, except for Mozambique (neutral when in the comparison were considered kinda Chunga), Namibia and Limpopo (neutral when in the comparison were considered kinda Kafue). Yellow Tanzania East shares more alleles with chacma than both kinda populations, with the exception of Zambia (which shares more alleles with kinda Chunga than Tanzania East), Namibia, and Cape (neutral when in the comparison there were kinda Chunga), Mozambique and museum specimens (neutral when in the comparison there were kinda Kafue) and Lesotho (neutral for both the comparisons). Very different were the results when comparing yellow Tanzania West and kinda populations. Yellow Tanzania populations share more alleles than kinda Kafue with all chacma populations except for Mozambique and the ancient specimens from Lesotho in which the imbalance was on the side of kinda. There are no differences in the allele sharing between yellow Tanzania West and kinda Chunga, with the exception of Zambia, which shares more alleles with kinda.

Based on these results, we attempted to characterise the extent of hybridisation between chacma populations and yellow/kinda baboons using f4 ratio tests (FIGURE 1.13-1.14-1.15). Since it is known that Zambia and Mozambique populations are certainly hybridised with yellow and/or kinda baboons, they were not used as external branches of the tree (A, B, or C) but only as test branches (X; please refer to FIGURE 0.6, in the introduction chapter). The other chacma populations (Cape, Limpopo, and Namibia), however, have been used as A, B, and X to best characterise which populations had the most presence of kinda and yellow in their genomes.

First, the proportion of one chacma population (B) present in another chacma population (X) was tested against the amount of a population from either kinda or yellow (C; FIGURE 1.13). In this case, the Cape, Limpopo, or Namibia populations were alternatively used as A, B, or X. These results suggest that Zambia and Mozambique have similar proportions of chacma ancestry when compared to each of the other chacma populations, with estimates around 50-60%. Limpopo shows higher chacma ancestry when compared to Zambia and Mozambique (around 80%). F4-ratio for Namibia in X is never supported, indicating that this population has a higher chacma component than Limpopo (B). Using Cape as B and Limpopo or Namibia as A, the same trend is observed, with Mozambique and Zambia having the lowest chacma component, followed by Limpopo (and the Transvaal Museum) and then Namibia.

Next, the presence of chacma in yellow populations was tested (FIGURE 1.14A). Kenya was excluded from this analysis since it is the most distant from chacma and has shown fewer signs of hybridisation compared to the other two yellow populations from Tanzania. Cape was used as A because it is the most distant from the hybridisation zone, while the other chacma

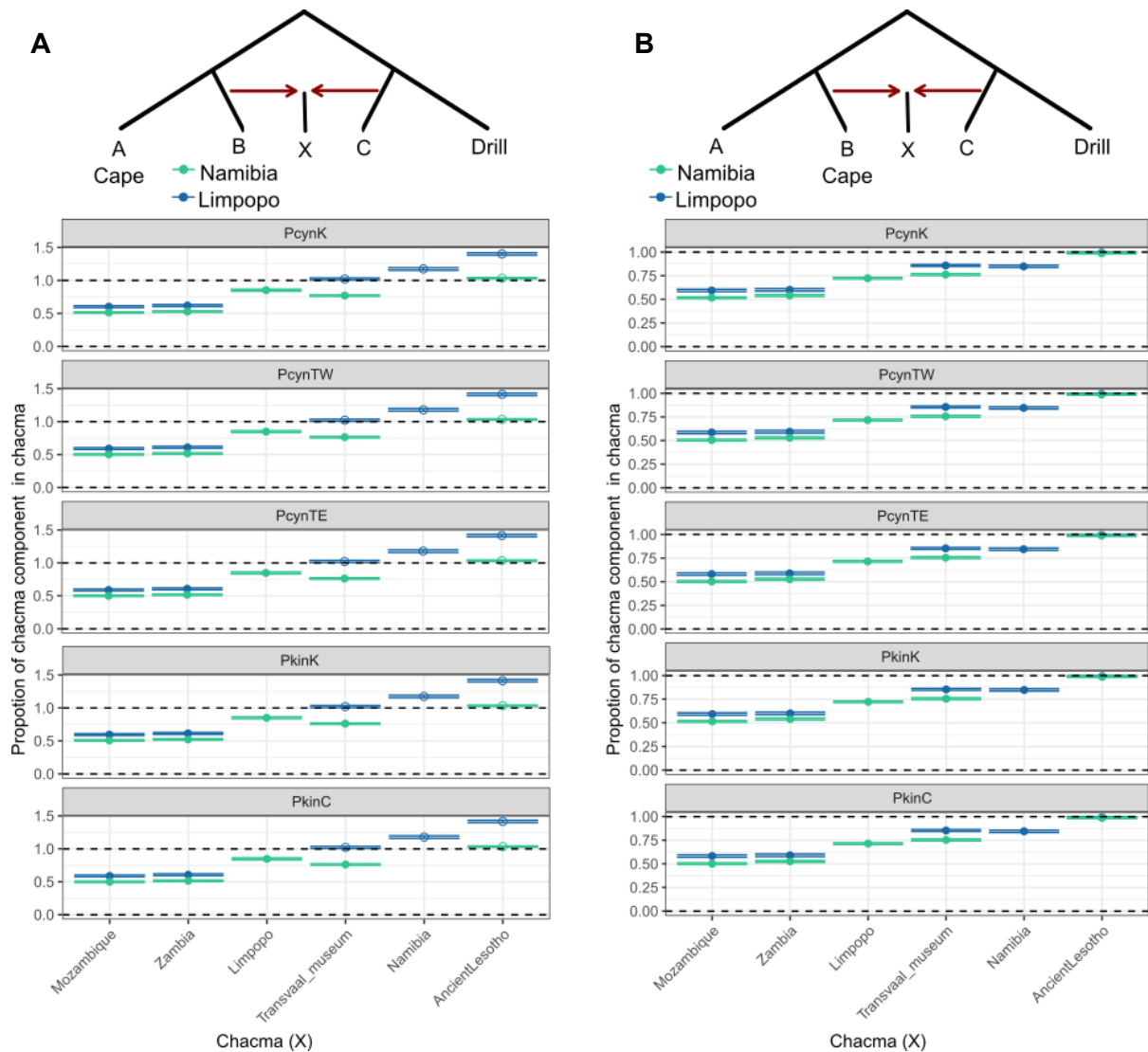


FIGURE 1.13 - Proportion of chacma component in chacma baboon populations: *f*₄-ratio tests for chacma populations using different tree topologies. The tree topologies (panel A and panel B) are indicated on the top. The empty dots refer to alpha values associated with unsupported Z values; the bars show the extension of three standard deviations. The name of the population above each box refers to the population used as branch C (PcynK is *P. cynocephalus* from Kenya, PcynTE is *P. cynocephalus* from East Tanzania, PcynTW is *P. cynocephalus* from West Tanzania, PkinK is *P. kindae* from Kafue, PkinC is *P. kindae* from Chunga). In panel A the colours refer to the chacma population used as branch B. In panel B the colours refer to the chacma population used as branch A.

populations were alternately used as B. The analysis revealed that both Eastern and Western Tanzanian yellow baboons have *f*₄-ratio values between 40% and 50% when tested with Mozambique and Zambia. In contrast, these values are between 20% and 30% when tested with other chacma populations as B, indicating that yellow baboons have traces of chacma ancestry that likely originate from Mozambique or Zambia-like populations.

The second step was to test the proportion of *P. cynocephalus* in chacma populations using Kenyan yellow as A and Cape as C, due to the reasons mentioned above (FIGURE 1.14B). When yellow East was used as B, the *f*₄-ratio values were around 10% for Zambia and Mozambique, while they were very close to 1 for the other chacma populations. This indicates that the frontier populations (Zambia and Mozambique) have a higher presence of yellow ancestry compared to the other chacma populations. when yellow West was used as B, none of the values

were supported. The f_4 -ratio values were all above 1 in these cases, but the same pattern described for yellow East was observed.

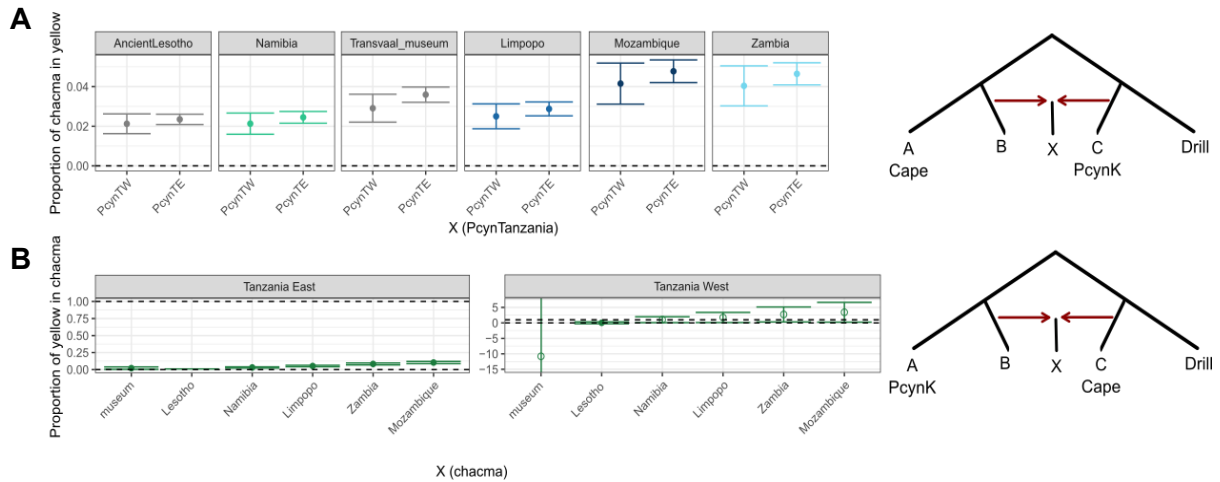


FIGURE 1.14 - f_4 -ratio tests for chacma and yellow baboon populations. A) testing the proportion of chacma component in yellow populations from Tanzania; B) testing the proportion of yellow component in chacma populations. In A and B, the tree topologies are indicated on the right. The empty dots refer to alpha values associated with unsupported Z values; the bars show the extension of three standard deviations. The name of the population above each box refers to the population used as branch B; PcynK refers to *P. cynocephalus* from Kenya. In panel A the colours refer to the chacma population used as branch B.

Moving on to kinda baboons, the percentage of chacma in the two kinda groups was tested (FIGURE 1.15A). First, the presence of chacma in kinda was tested using the Cape population in A, while in B, all other chacma populations were alternately tested, and in C, the alternative kinda population to the one being tested was used. The results showed that when kinda Kafue was used as the test population (X), the f_4 ratio values were never supported, indicating that this population likely has less chacma ancestry than the Chunga population. In fact, when the latter population was used as the test, all values were supported. Moreover, it was observed that the values were all very close to zero, but when Mozambique and Zambia were used as B, slightly higher values, around 3%, were obtained.

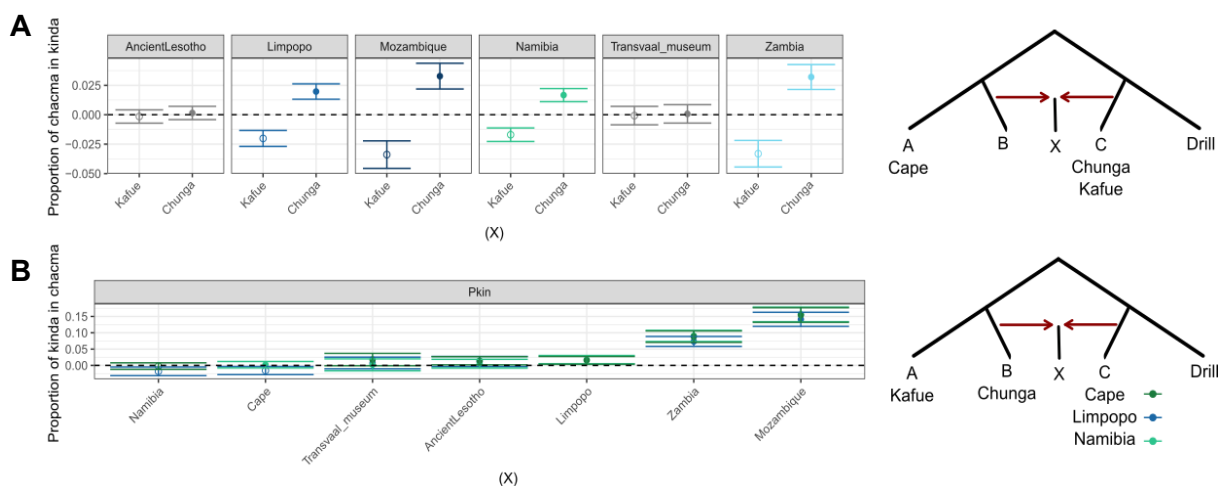


FIGURE 1.15 - f_4 -ratio tests for chacma and kinda baboon populations. A) testing the proportion of chacma component in kinda populations; B) testing the proportion of kinda component in chacma populations. In A and B, the tree topologies are indicated on the right. The empty dots refer to alpha values associated with unsupported Z values; the bars show the extension of three standard deviations. The name of the population above each box refers to the population used as branch B; Pkin in panel B refers to *P. kindae* from Chunga. In panel A the colours refer to the chacma population used as branch B. In panel B the colours refer to the chacma population used as branch C.

Subsequently, the presence of kinda (A and B) in each chacma population (X) was tested (FIGURE 1.15B). In this case, kinda Kafue was used in A and kinda Chunga in B, as Kafue had been shown to have less chacma than Chunga. In C, Limpopo, Cape, and Namibia were alternately used, while in X, all chacma populations were tested. The results showed that Zambia and Mozambique consistently had the highest f4 ratio values, ranging between 10% and 15%, while all other populations showed zero values. Only Limpopo had positive values, albeit very close to zero, indicating that this population has slightly more kinda ancestry than Namibia and the Cape.

1.3.4 - Demographic History

The reconstruction of demographic changes across time was performed using the PSMC software (Li & Durbin, 2011). All *P. ursinus* samples showed a decrease in coalescent curves when comparing more recent and ancient estimates (FIGURE 1.16A). However, two major patterns can be highlighted over the last 200,000 years. While samples in the Northern part of the chacma range (Zambia and Mozambique) maintained a large population size for longer times, the other *P. ursinus* populations (including the Cape, Limpopo and Namibia) experienced a steadier decrease in effective population size starting 500k years ago.

Comparing these results with the yellow and kinda PSMC results, a significantly lower trend can be observed in chacma populations (FIGURE 1.16B). Kinda and yellow populations have much higher curves (yellow Tanzanian West have the higher ones), with a decrease in the last 20k years, reaching the curves of the chacma populations.

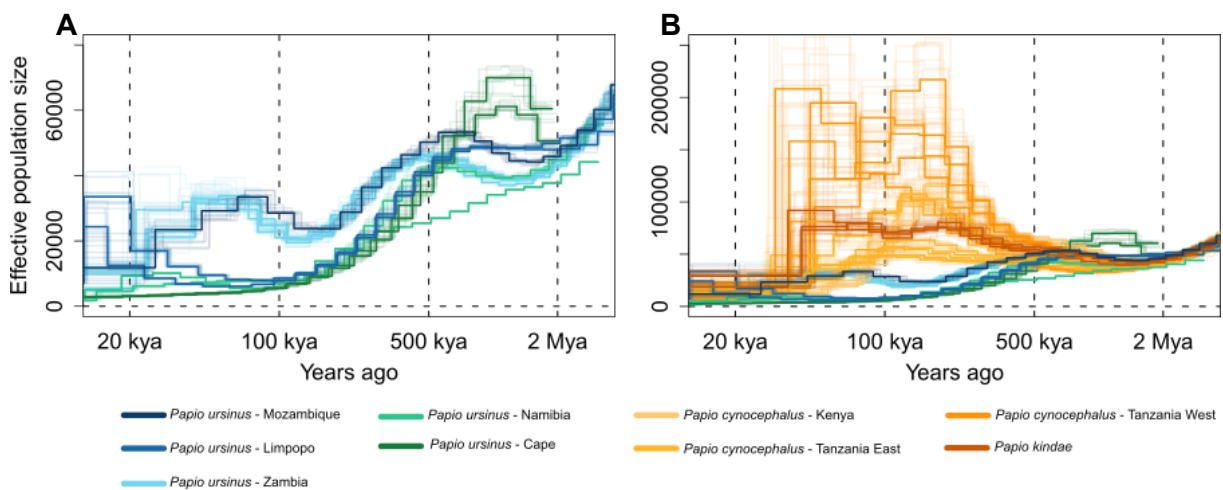


FIGURE 1.16 - Demographic history of southern *Papio* populations. Changes in effective population size across time estimated using PSMC A) for only chacma populations; B) for all yellow, kinda and chacma populations. X-axis on logarithmic scale.

To estimate the timing of the most recent gene flow events between different chacma populations (FIGURE 1.17A-E), we reconstructed the demographic curves of artificial F1 hybrids of chromosome X using hPSMC (Cahill et al. 2016). According to coalescent theory, the lack of coalescent events suggests extremely large populations, leading to curves that increase exponentially at the time of the most recent common ancestor. The curves for the hybrids between

all chacma populations show this pattern at 5-10 thousand years ago (kya), suggesting this was the time of the last genetic exchange (with the most ancient between Mozambique and Cape; FIGURE 1.17A). There are only three exceptions to this pattern: the curves for the hybrids Limpopo-Cape (FIGURE 1.17DE) and Limpopo-Zambia (FIGURE 1.17BD), the related curves suggesting a last genetic exchange more recent than 5kya and the curve for hybrid Zambia-Mozambique (FIGURE 1.17AB) which suggests ongoing genetic exchange.

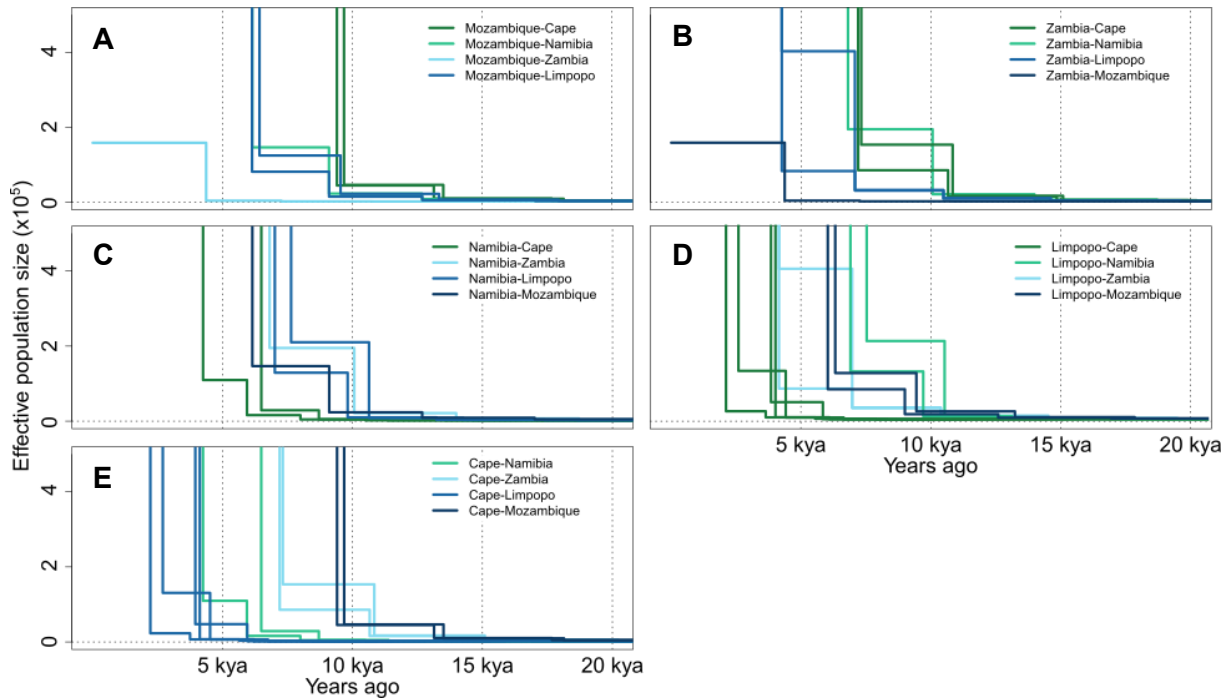


FIGURE 1.17 – Last gene flow timing between chacma baboon populations. hPMSC coalescence curves of artificial X chromosome hybrids between A) the Mozambique population and the other chacma populations; B) the Zambia population and the other chacma populations; C) the Namibia population and the other chacma populations; D) the Limpopo population and the other chacma populations; E) the Cape population and the other chacma populations.

The same approach was applied to observe the timing of the most recent gene flow between chacma and other populations in Southern Africa, the yellow and kinda baboons (FIGURE 1.18AB). The curves for hybrids between Zambia and Mozambique and the neighbouring *Papio* forms suggest more recent events if compared with the curves for hybrids between other chacma and the neighbour forms. In particular, for the chacma populations in the North (Zambia and Mozambique), the curves suggest that the last contacts were not older than 30kya, while for the other chacma populations, the last most recent contacts with yellow populations were at least 40-50kya.

By closely examining the curves of hybrids between yellow and chacma populations (FIGURE 1.18A), we can note that for the Mozambican population, the last contact with Mikumi (Tanzania East), Mahale (Tanzania West) and Amboseli (Kenya) populations was at 15kya, while with all the others yellow populations was around 20kya. For the Zambian population, the last genetic exchange with Mahale and Katavi (Tanzania West) was 10kya, with Mikumi and Amboseli 15kya and with all the other yellow populations between 20 and 30kya. For the Cape population, the last gene flow with Mahale and Katavi was 40kya, with Ruaha, Mikumi (Tanzania East) and Amboseli was 50kya, with Udzungwa, Selous (Tanzania East) and Issa Valley

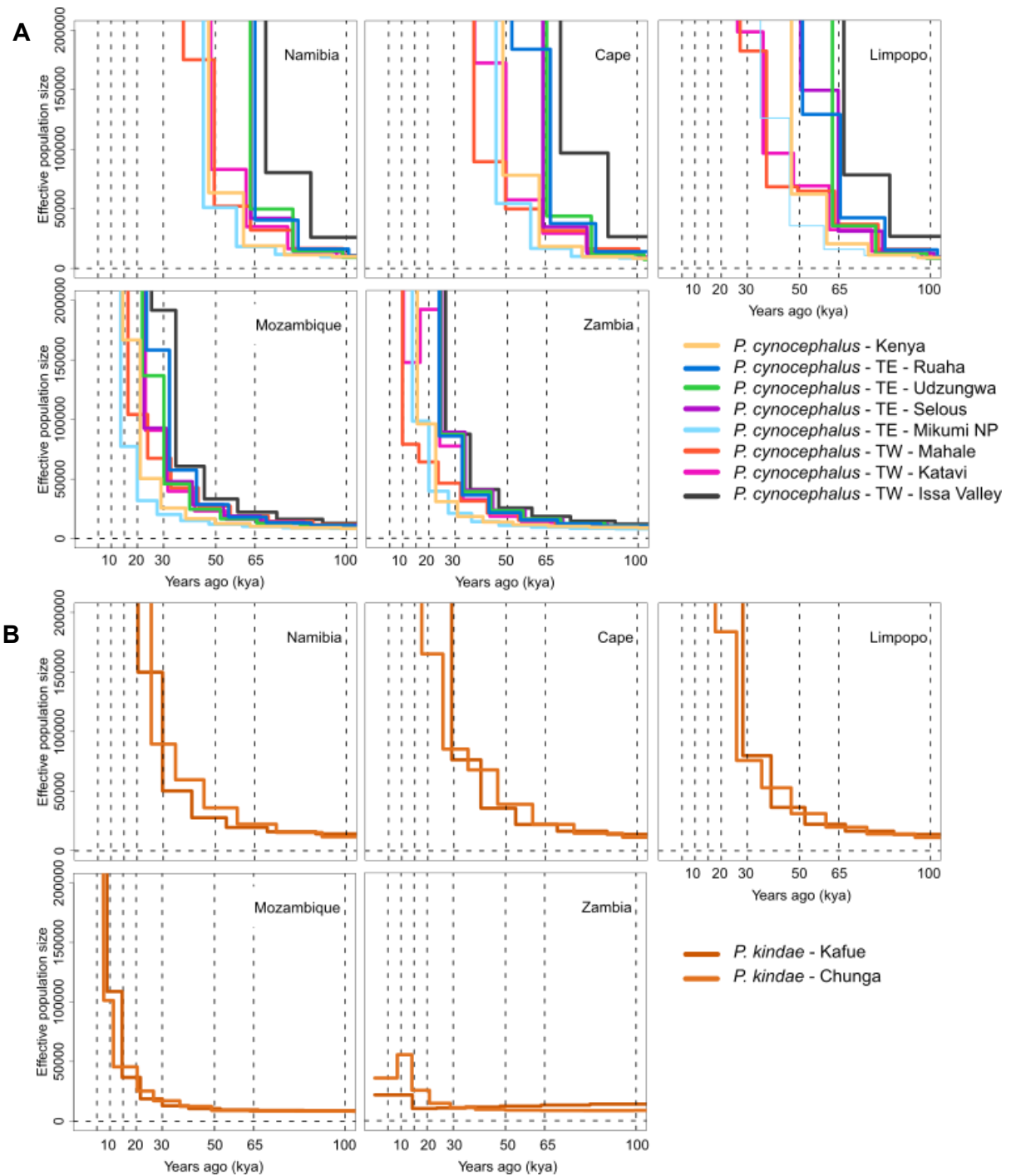


FIGURE 1.18 - Last gene flow timing in southern baboons. hPMSMC coalescence curves of artificial *X* chromosome hybrids between A) chacma and yellow populations; B) chacma and kinda populations. In A and B each box refers to hPMSMC analyses on artificial hybrids between the chacma population indicated in the corresponding box and the yellow or kinda population associated with the corresponding colour.

populations 65kya. For the Namibian population, the last contact with Mahale, Katavi, Amboseli and Mikumi was between 40 and 50kya, while with all the other yellow populations it was 65kya. Regarding the population of Limpopo, it can be noted that the curves that suggest the most recent contacts with the Mahale and Katavi populations between 20 and 40kya, with the Amboseli 50kya, with the Mikumi between 30 and 50kya, the Ruha, Selous, Udzungwa and Issa Valley populations between 50kya and 65kya.

The pattern for the last interaction between chacma and kinda is different (FIGURE 1.18B). All chacma populations, show gene exchange more recent than the last contacts with yellow, with no differences for kinda from Chunga and Kafue. For the kinda-Zambia hybrid, the curve suggests ongoing genetic exchange, the curve for the Mozambique-kinda hybrid suggests a last contact 10kya, while for the other chacma the last gene flow is supposed to be between 15 and 30kya.

1.4 – Discussion

The results of the different analyses reveal a complex and nuanced picture for the population structure, gene flow, and demographic history of *Papio ursinus* and neighbouring baboon species (*P. cynocephalus* and *P. kindae*) in Southern Africa.

1.4.1 – *P. ursinus* Populations at the Northern Edge of the Species Range: Zambia and Mozambique.

The patterns of shared alleles highlighted by the analysis of the D-statistics suggest that both of the examined *P. ursinus* populations at the Northern edge of the species' range (Zambia and Mozambique) have been exposed to *P. cynocephalus* introgression. The source of this introgression likely comes from a population more closely related to the Eastern *P. cynocephalus* tested here than the Western one. However, the Zambian *P. ursinus* appear to harbour an additional genetic component, related to *P. kindae*, as previously suggested (Sørensen et al. 2023). A similar three-way admixture has also been proposed for Western *P. cynocephalus* between kinda, yellow and olive (Sørensen et al. 2023), which could explain why this population appears more closely related to Zambian than Mozambican baboons.

The similar affinity with Eastern yellow baboons, coupled with the differential association with *P. kindae*, suggests that Northern chacma populations may have experienced distinct dynamics in their interactions with nearby *Papio* species. It is possible, for instance, that the observed contribution of Western Tanzanian yellow baboons to Zambian populations arrived via *P. kindae*, while this contribution did not reach Mozambique, as the likely source of hybridisation there was a population close to Eastern Tanzanian yellow baboons. Moreover, these events appear to be temporally distinct. The hPSMC analysis indicates that hybridisation with *P. kindae* in Zambia seems to be ongoing, whereas in Mozambique, while the contact with the yellow baboons tested was relatively recent, it no longer appears to be occurring. For both populations, the hPSMC curves suggest that the last contact with yellow baboon dates back to between 20 and 30 kya.

These findings are not unexpected, given that hybridisation between chacma and kinda baboons has been observed and studied in the Kafue National Park region (Burrell 2009; C. J. Jolly et al. 2011; Chiou et al. 2021), whereas no hybridisation has been observed in the field in Gorongosa, Mozambique. However, there are signs of ancient hybridisation events between yellow and chacma baboons in modern individuals (Martinez et al. 2019; Santander et al. 2022). These results provide further support for the notion that the genetic barriers between these baboon

forms are permeable and not strong enough to prevent successful interbreeding, thus challenging the axiom of the biological species concept.

Additionally, we note that the newly added genomic samples from Gorongosa Park are consistent with previous results, showing evidence of introgression from *P. cynocephalus* (Santander et al. 2022). This confirms that introgression is a defining feature of the local population. Given the lack, or minimal presence, of population structure within the park and nearby regions (e.g., Catapù; Ferreira da Silva et al. submitted), further genome sampling is expected to provide the context necessary for understanding the evolutionary dynamics that have shaped the genome-wide distribution of these introgression signatures. Moreover, it will help characterise their phenotypic consequences, including adaptations to the highly heterogeneous Gorongosa ecosystem (Carvalho et al. submitted). Finally, it is worth noting that Gorongosa hosts two other primate species, the vervet and samango monkeys (genera *Chlorocebus* and *Cercopithecus*, respectively; Santander et al. 2022; Gaynor et al. 2021). Their genomic and genetic analysis is expected to clarify to what extent central Mozambique has historically functioned as a crossroads for interactions with nearby populations and species.

A more detailed analysis of the Zambian population reveals significant hybridisation between kinda and chacma baboons. In the PCA, many individuals fall along a continuum between the two groups, with ADMIXTURE confirming a genomic mosaic between these species. This hybridisation gradient suggests that species boundaries are not rigid and that individuals retain varying proportions of kinda and chacma genomes. The use of the f_4 -ratio test provided further insight into the level of kinda genomic contribution to Zambian chacma baboons. Interestingly, the Zambian population exhibits notable variability in the proportions of *P. kindae* genomes, indicating highly localised hybridisation dynamics. The fact that using external populations (Namibia, Limpopo) as a reference for gene flow reduces the *P. kindae* contribution suggests that hybridisation is a phenomenon specific to the Zambian context, where it is more pronounced.

1.4.2 - Hybridisations Events and Population Structure

The presence of traces of kinda and yellow baboon introgression in chacma genomes is not limited to the two Northern populations but also extends to other chacma populations under consideration. However, the signals are significantly weaker and seem to fade with increasing distance from the Northern hybridisation hotspots. Notably, the Southernmost population, in the Cape region, exhibits minimal signals of gene flow with the other two species, while the two populations located midway along the longitudinal range show intermediate signals. Additionally, the Namibian population presents weaker signals compared to Limpopo, yet reflects the patterns observed in the neighbouring Zambian population. The Limpopo population, on the other hand, displays intermediate patterns between Zambia and Mozambique, which mirrors its geographically equidistant position between the two. In the PCA, the Limpopo population (along with a museum specimen from a nearby geographical area) is positioned between the two Northern populations and those from the South and West (Cape and Namibia). This intermediate pattern is also observed in the hPSMC, where the curves of artificial hybrids with kinda and yellow baboons shift towards more recent times compared to the Cape and Namibia, but more ancient than Mozambique and Zambia.

The relationships between chacma populations appear to be much more complex, with gene flow interruptions occurring very recently (between 5,000 and 10,000 years ago), and in some cases, still ongoing (Zambia and Mozambique). The patterns of allele sharing suggest stronger links between Northern populations with each other, and between Southern populations with each other, with Limpopo once again in an intermediate position, maintaining connections with both the North and the South. Overall, a major differentiation in population structure is observed, driven by differential hybridisation with neighbouring species, suggesting that the insertion of the Northern chacma haplogroup into the branches of kinda and Southern yellow, generating a paraphyly within chacma baboons, is likely due to nuclear swamping following hybridisation (Zinner et al. 2013; Roos et al. 2021; Kopp et al. 2023).

Demographic curves for Zambia and Mozambique suggest a higher effective population size compared to other chacma populations, which may be due to gene flow with neighbouring species and/or populations, as well as potential ecological differences in habitat suitability (Stone et al. 2016; Sithaldeen, Ackermann and Bishop 2015). Estimates of heterozygosity reveal clear differences in genetic diversity among chacma populations. Chacma baboons from the Cape, Namibia, and Limpopo exhibit the lowest levels of heterozygosity, consistent with their more isolated and genetically homogeneous populations. In contrast, chacma baboons from Mozambique and Zambia show higher heterozygosity, likely due to hybridisation with kinda and yellow baboons, which has probably introduced new genetic variation in these populations. Among other species, yellow baboons from Western Tanzania display the highest heterozygosity, possibly the result of gene flow and admixture with neighbouring species (e.g., olive baboons; Sørensen et al. 2023).

Both ADMIXTURE and PCA analyses indicate that Zambian and Mozambican populations exhibit distinct differences compared to other chacma populations, likely due to a higher contribution from *P. kindae* and *P. cynocephalus*. In the PCA, hybrids from the hybridisation region form a continuum between Northern chacma and kinda, suggesting a blending of forms. This gradient potentially extends from the Cape population, indicating that as more intermediate populations are sampled between the Cape and Zambia, a continuum of individuals could potentially emerge. This observation aligns with Jolly's (1993) description of subspecies, where he noted a certain continuity between forms. Indeed, if one compares an individual from the Northern range to one from the South, there are substantial differences, with Southern individuals being darker and larger, while Northern individuals are lighter and smaller. However, individuals living in the middle of these extremes display intermediate characteristics, which led Jolly to propose a hypothetical additional subspecies that bridges the Northern and Southern forms. A similar continuum of forms has been observed between chacma and Southern yellow baboons (Martinez et al. 2019), and it may extend even further between Northern yellow baboons and Eastern olive baboons (Jolly 1993). In short, these hybridisation events in baboons blur the boundaries between different forms, making it difficult to pinpoint where one form ends and another begins.

Based on these findings, it seems plausible that the kinda contribution reached the Mozambican population through gene flow from Zambia-like populations, while the yellow baboon contribution reached Zambia both through Mozambique-like populations and via kinda. In the former case, it is possible that yellow baboons arrived from Eastern Tanzania, while in the latter, they likely arrived from Western Tanzania. These contributions were then transmitted to other

chacma populations through gene flow and exchanges between populations, progressively diluting these contributions until the Cape population, where the lowest signals of allele sharing with kinda and yellow baboons are recorded. This gene flow has shaped the differentiation and structure observed in chacma populations today.

The study of baboons from Luangwa Valley in Zambia could further clarify the relationships between the three species under the PSC in the future, as that region appears to be a potential hybridisation hotspot for the three forms (Koop et al. 2023).

1.5 - Conclusions

These findings provide important insights into the evolutionary history of baboon populations in Southern Africa. The complex patterns of gene flow between chacma, kinda and yellow baboons, as well as the significant population structure among chacma populations, illustrate the dynamic nature of baboon evolution in this region. The study also highlights the importance of hybridisation in shaping the genetic landscape of baboons, particularly in hybrid zones like Zambia and Mozambique, where multiple species have come into contact.

The results also emphasize the role of geographic isolation in shaping genetic diversity and population structure, with Southern populations (Cape) being more isolated and genetically homogenous, while Northern populations (Zambia, Mozambique) are more diverse due to ongoing gene flow with neighbouring species. Overall, this study contributes to our understanding of the complex demographic history and evolutionary processes that have shaped the baboon populations of Southern Africa and highlights the importance of hybridisation in the evolution of primates.

1.6 - References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655–1664.
<https://doi.org/10.1101/gr.094052.109>

Allan, A. T. L., Bailey, A. L., and Hill, R. A. (2020). Habituation is not neutral or equal: Individual differences in tolerance suggest an overlooked personality trait. *Science advances*, 6(28), eaaz0870. <https://doi.org/10.1126/sciadv.aaz0870>

Batra, S. S., Levy-Sakin, M., Robinson, J., Guillory, J., Durinck, S., Vilgalys, T. P., Kwok, P. Y., Cox, L. A., Seshagiri, S., Song, Y. S., and Wall, J. D. (2020). Accurate assembly of the olive baboon (*Papio anubis*) genome using long-read and Hi-C data. *GigaScience*, 9(12), giaa134.
<https://doi.org/10.1093/gigascience/giaa134>

Burrell, A. S. (2009). *Phylogenetics and Population Genetics of Central African Baboons*. New York University.

Cahill, J. A., Soares, A. E., Green, R. E., and Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1699), 20150138. <https://doi.org/10.1098/rstb.2015.0138>

Carvalho, S., Farassi, R., Mathe, M., and Bobe, R. (submitted). Introduction to the Issue: Primates in highly seasonal environments: the case of Gorongosa National Park.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>

Chiou, K. L., and Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific reports*, 8(1), 1975. <https://doi.org/10.1038/s41598-018-20427-9>

Chiou, K. L., Bergey, C. M., Burrell, A. S., Disotell, T. R., Rogers, J., Jolly, C. J., & Phillips-Conroy, J. E. (2021). Genome-wide ancestry and introgression in a Zambian baboon hybrid zone. *Molecular ecology*, 30(8), 1907–1920. <https://doi.org/10.1111/mec.15858>

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>

Ferreira da Silva, M. J., Tralma, P., Colmonero-Costeira, I., Mota, M., Farassi, R., Hammond, P., Lewis-Bevan, L., Bamford, M. K., Biro, D., Lüdecke, Y. Mathe, J., Bobe, R., Capelli, C., Carvalho, S., and Martínez, F. I. (accepted). Sex-mediated gene flow of grayfoot baboons (*Papio ursinus griseipes*) in a highly seasonal habitat of Gorongosa National Park. *International Journal of Primatology*. Preprint: <https://www.researchsquare.com/article/rs-4616083/v1>

Gaynor, K. M., Daskin, J. H., Rich, L. N., and Brashares, J. S. (2021). Postwar Wildlife Recovery in an African Savanna: Evaluating Patterns and Drivers of Species Occupancy and Richness. *Animal Conservation* 24 (3): 510–22.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C., & Rogers, J. (2011). Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river valley, Zambia. *American journal of primatology*, 73(3), 291–303. <https://doi.org/10.1002/ajp.20896>

Jolly, C. J. (1993). Species, Subspecies, and Baboon Systematics. In *Species, Species Concepts and Primate Evolution*, 67–108. *Springer US*, Boston.

Kopp, G. H., Sithaldeen, R., Trede, F., Grathwol, F., Roos, C., & Zinner, D. (2023). A Comprehensive Overview of Baboon Phylogenetic History. *Genes*, 14(3), 614. <https://doi.org/10.3390/genes14030614>

Kingdon, J. (1997). The Kingdon Field Guide to African Mammals. *Academic Press*, London.

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Kharitanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., and Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature communications*, 9(1), 5018. <https://doi.org/10.1038/s41467-018-07483-5>

Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>

Martina, C., Cowlshaw, G., & Carter, A. J. (2020). Exploring individual variation in associative learning abilities through an operant conditioning task in wild baboons. *PloS one*, 15(4), e0230810. <https://doi.org/10.1371/journal.pone.0230810>

Martinez, F. I., Capelli, C., Ferreira da Silva, M. J., Aldeias, V., Alemseged, Z., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Habermann, J. M., Lüdecke, T., Madiquida, H., Mathe, J., Negash, E., Paulo, L. M., Pinto, M., Stalmans, M., Tátá, F., & Carvalho, S. (2019). A missing piece of the Papio puzzle: Gorongosa baboon phenostructure and intragenetic relationships. *Journal of human evolution*, 130, 1–20. <https://doi.org/10.1016/j.jhevol.2019.01.007>

Mathieson, I., Abascal, F., Vinner, L., Skoglund, P., Pomilla, C., Mitchell, P., Arthur, C., Gurdasani, D., Willerslev, E., Sandhu, M. S., & Dewar, G. (2020). An Ancient Baboon Genome Demonstrates Long-Term Population Continuity in Southern Africa. *Genome biology and evolution*, 12(4), 407–412. <https://doi.org/10.1093/gbe/evaa019>

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>

Miquel, C., Bellemain, E., Poillot, C., Bessière, J., Durand, A., and Taberlet, P. (2006). Quality Indexes to Assess the Reliability of Genotypes in Studies Using Noninvasive Sampling and Multiple-tube Approach. *Molecular Ecology Notes* 6 (4): 985–88.

Mutti, G., Oteo-Garcia, G., Caldon, M., da Silva, M. J. F., Minhós, T., Cowlshaw, G., Gottelli, D., Huchard, E., Carter, A., Martinez, F. I., Raveane, A., & Capelli, C. (2023). Assessing the recovery of Y chromosome microsatellites with population genomic data using Papio and Theropithecus genomes. *Scientific reports*, 13(1), 13839. <https://doi.org/10.1038/s41598-023-40931-x>

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>

Robinson, J. A., Belsare, S., Birnbaum, S., Newman, D. E., Chan, J., Glenn, J. P., Ferguson, B., Cox, L. A., & Wall, J. D. (2019). Analysis of 100 high-coverage genomes from a pedigreed captive baboon colony. *Genome research*, 29(5), 848–856. <https://doi.org/10.1101/gr.247122.118>

Rogers, J., Raveendran, M., Harris, R. A., Mailund, T., Leppälä, K., Athanasiadis, G., Schierup, M. H., Cheng, J., Munch, K., Walker, J. A., Konkel, M. K., Jordan, V., Steely, C. J., Beckstrom, T. O., Bergey, C., Burrell, A., Schrempf, D., Noll, A., Kothe, M., Kopp, G. H., ... Baboon Genome Analysis Consortium (2019). The comparative genomics and complex population history of *Papio* baboons. *Science advances*, 5(1), eaau6947. <https://doi.org/10.1126/sciadv.aau6947>

Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American journal of physical anthropology*, 174(3), 407–417. <https://doi.org/10.1002/ajpa.24186>

Santander, C., Molinaro, L., Mutti, G., Martínez, F. I., Mathe, J., Ferreira da Silva, M. J., Caldon, M., Oteo-Garcia, G., Aldeias, V., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Hammond, P., Lüdecke, T., Pinto, M. J., Meira Paulo, L., Stalmans, M., Regala, F. T., ... Capelli, C. (2022). Genomic variation in baboons from central Mozambique unveils complex evolutionary relationships with other *Papio* species. *BMC ecology and evolution*, 22(1), 44. <https://doi.org/10.1186/s12862-022-01999-7>

Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., Alessandri, L., Arienzo, I., De Angelis, F., Rolfo, M. F., Skeates, R., Silvestri, L., Beckett, J., Talamo, S., Dolfini, A., Miari, M., ... Scheib, C. L. (2021). Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Current biology : CB*, 31(12), 2576–2591.e12. <https://doi.org/10.1016/j.cub.2021.04.022>

Sithaldeen, R., Bishop, J. M., & Ackermann, R. R. (2009). Mitochondrial DNA analysis reveals Plio-Pleistocene diversification within the chacma baboon. *Molecular phylogenetics and evolution*, 53(3), 1042–1048. <https://doi.org/10.1016/j.ympev.2009.07.038>

Sithaldeen, R., Ackermann, R. R., & Bishop, J. M. (2015). Pleistocene aridification cycles shaped the contemporary genetic architecture of Southern African baboons. *PloS one*, 10(5), e0123207. <https://doi.org/10.1371/journal.pone.0123207>

Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., Storer, J. M., Kuhlwilm, M., Fontserè, C., Seshadri, L., Bergey, C. M., Burrell, A. S., Bergman, J., Phillips-Conroy, J. E., Shiferaw, F., Chiou, K. L., Chuma, I. S., Keyyu, J. D., Fischer, J., Gingras, M. C., ... Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science (New York, N.Y.)*, 380(6648), eabn8153. <https://doi.org/10.1126/science.abn8153>

Stone, O. M. L., Herries, A. I. R., Brink, J. S., and Laffan, S. W. (2016). The Chacma Baboon (*Papio Ursinus*) through Time: A Model of Potential Core Habitat Regions during a Glacial–interglacial Cycle. *Evolutionary Ecology* 30 (4): 755–82. <https://doi.org/10.1007/s10682-016-9833-8>

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, 2019, pp. 314-324. <https://doi.org/10.1109/IPDPS.2019.00041>

Wall, J. D., Schlebusch, S. A., Alberts, S. C., Cox, L. A., Snyder-Mackler, N., Nevonen, K. A., Carbone, L., & Tung, J. (2016). Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of Admixture in wild baboons. *Molecular ecology*, 25(14), 3469–3483. <https://doi.org/10.1111/mec.13684>

Wilson Sayres, M. A., Venditti, C., Pagel, M., & Makova, K. D. (2011). Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution; international journal of organic evolution*, 65(10), 2800–2815. <https://doi.org/10.1111/j.1558-5646.2011.01337.x>

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American journal of physical anthropology*, 150(1), 133–140. <https://doi.org/10.1002/ajpa.22185>

Chapter 2 - Hybridisation in Eastern Africa: *P. anubis* and *P. hamadryas*

The material presented in this chapter refers to a manuscript in preparation: Caldon et al. “Museomics of Eastern African baboons provide insights on their population structure”.

Preliminary results from this chapter were presented at the SBE (Society for Molecular Biology and Evolution) satellite meeting focused on ancient DNA which was held at Trinity College in Dublin, Ireland, in July 2024.

2.1 – Introduction

2.1.1 – Science and Colonialism

In the XIX century, scientific expeditions were crucial for the advancement of biological sciences. The specimens collected from around the globe reshaped fundamental concepts in biology and stimulated new research in comparative anatomy, biogeography, and adaptations. These journeys, both overland and by sea, played a crucial role in the expansion of European colonialism (Browne 1992). Every scientific observation was viewed as a contribution to colonisation, and the collection of specimens became a highly valued national activity. Expeditions, often organised by governments, involved scientists as guests or temporary officers aboard ships or under military supervision for land expeditions (Browne 1992). This state-supported model persisted throughout the XIX and much of the XX century. For example, in 1831, Charles Darwin's voyage on HMS Beagle aimed to map the South American coastline, the Falkland Islands, and the Galápagos Islands to bolster British control over these regions. Such colonial scientific expeditions frequently led to the subjugation of native peoples, slavery, and genocide (Das and Lowe 2018).

The specimens collected during those initiatives were vital for studies in classification, comparative anatomy, biogeography, and evolutionary ecology. From the mid-century onwards, many European governments established natural history museums to preserve and study these collections, with the most prominent ones located in Paris, London, Vienna, and Berlin. These centres became crucial for processing and studying specimens (Schildkrout 1989).

The imperialistic notion of 'centre and periphery' suited European and North American naturalists, who saw the centre - with its vast intellectual community and specimen collections - as the most suitable place to turn natural objects into scientific 'facts'. From this perspective, these expeditions and the knowledge they generated represented a form of exploitation. Collectors rarely showed concern for biological conservation. As a result, specimens displayed in Western metropolitan museums became symbols of both the production of scientific knowledge and imperialistic expansion (Miriti, Rawson, and Mansfield 2023; Browne 1992).

In the latter half of the XIX century, Africa, previously limited to coastal European presence, became a prime target for colonial expansion. European powers engaged in a race to acquire new territories and subjugate indigenous populations. Within fifty years, seven European countries - Britain, France, Spain, Portugal, Germany, Italy, and Belgium - had conquered almost the entire African continent. This period saw a surge in expeditions to explore and map the continent's interior, particularly its major rivers. These expeditions brought a vast amount of scientific material to European cities, much of which remains on display in museums today. The history of modern natural science is so intricately linked with European colonialism that they are virtually inseparable (Das and Lowe 2018; Miriti, Rawson, and Mansfield 2023).

Like the other European nations, Italy participated in the scramble for colonies during the late XIX and early XX centuries. Italian explorers and scientists were involved in expeditions to various parts of Africa and other regions, contributing to the broader European colonial enterprise. Italy's colonial ambitions were evident in its acquisitions of territories such as Eritrea, Somalia, and Libya. These colonial pursuits were accompanied by scientific expeditions aimed at mapping and documenting the newly acquired lands, often to enhance Italy's strategic and economic interests (Podestà 1998).

In 1869, the Suez Canal was opened, and the president, a vice president, and the secretary-general of the Italian Geographical Society attended the inauguration. The secretary-general, Orazio Antinori, while he was in Egypt, came into contact with a group of Italian emigrants

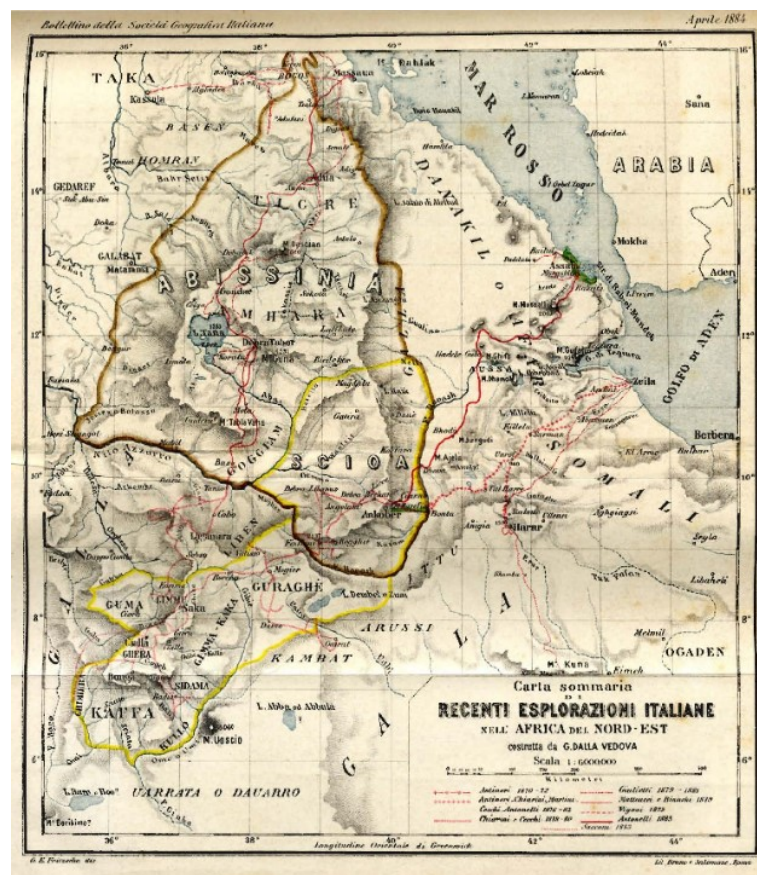


FIGURE 2.1 – Geographical map of Italian explorations in Eastern Africa (1870-1883). From *Bolletino della Società Geografica Italiana*, Aprile 1884.

who were privately organising an agricultural enterprise in Eritrea and attempted to obtain government support for the initiative. To strengthen the request, the Geographical Society organised its first expedition (1870), led by its secretary (FIGURE 2.1). During this expedition, Antinori witnessed the signing of the lease contract for the Bay of Assab on the Danakil coast (the first Italian base in Africa) on behalf of the Italian government. The expedition to the Bay of Assab and the land of the Bogos in Ethiopia in 1870 laid the political foundations for the Italian colonisation of East Africa, and it brought back to Italy the first African specimens to be preserved in the Natural History Museum of Genoa (Cerretti 2003; Doria and Poggi 2017). Most of the specimens used in this Chapter were obtained from a few Italian expeditions in Eastern Africa.

Marquess Orazio Antinori (1811-1882), as mentioned, was the secretary and one of the founders of the Italian Geographical Society, and on its behalf, he undertook numerous expeditions to the African continent. In 1876, he led the "Great Expedition to the Equatorial Lakes," which ascended the Awash River, crossed the territories of the Somali Issa and Afar, and reached Shewa (FIGURE 2.1), where he met Negus (King) Menelik II. The Negus offered him land in Let Marefià (Ankober), where he founded the first Italian geographical health station in Africa, from which numerous expeditions were launched and where he lived for the rest of his life. The station at Let Marefià was later directed by Vincenzo Ragazzi (1856-1929) between 1885 and 1890, during which time he collected a substantial number of zoological specimens, including many baboons (Doria and Poggi 2017).

Antinori and Ragazzi imported most of the specimens used in this work now conserved in the Genoa Museum. Another late 19th-century expedition that brought the material (that was used in this work) to the Genoa Museum was the 1891 expedition to East Africa led by Eugenio Ruspoli (1866-1893). This expedition departed from Berbera town in Somalia, crossed the Ogaden (Ethiopia), reached the Uebi Shebelle River, and then returned to Berbera (Doria and Poggi 2017).

Other specimens used was collected by Raimondo Franchetti (1889-1935), who undertook numerous expeditions in East Africa, notably organising an expedition to Danakil in 1928 that concluded in Assab. Additional specimens came from Saverio Patrizi (1902-1957)'s 1927 expedition, which explored the Uele River (in the ex-Belgian Congo, now the Democratic Republic of Congo), reaching the Nile Valley; and from his 1934 expedition on behalf of the Genoa Museum to explore the Somali regions of Juba and Oltre Giuba. Many samples from the Genoa Museum were collected by Edoardo Zavattari (1883-1972), known for his theories on biological racism, the founding of the "Rivista di Biologia Coloniale" (Journal of Colonial Biology) in 1938, and his endorsement of the Manifesto of Racist Scientists. The expedition that brought the specimens used in this work to Italy was his 1939 expedition through Southern Ethiopia, which brought him to the region of Borana and Segan-Omo populations (Doria and Poggi 2017). The most recent specimen included in this work comes from the Museum of Milan and was imported by Giuseppe Scortecci (1898-1973) in 1953 when he returned to Somalia after the Second World War.

2.1.2 - Museomics

Biological museum material, although often imported into European (and US) museums in ways that were not always ethical, directly or indirectly causing harm to local populations

(Ashby and Machin 2021), today more than ever becomes an invaluable source for now-days science thanks to the revolution of ancient DNA, which has led to the emergence of a new field of study, museomics.

Museomics (the union of the two words “museum” and “genomics”) is an emerging approach focused on the genomic analysis of specimens conserved within museum collections (Lalueza-Fox 2022). This field presents the same challenges encountered in ancient DNA studies, involving the use of genomic data that is partially degraded and incomplete. Therefore, it is necessary to process them differently from modern genomes, employing methodological adjustments, especially concerning the calling of nucleotide variants (Lalueza-Fox 2022; Raxworthy and Smith 2021; Wandeler et al. 2007). However, it has the advantage of investigating historical populations that may be locally extinct, thus providing a better understanding of the evolutionary history of a species. Additionally, it is a very useful approach when field sampling is difficult or impossible, for example, in areas that are hard to access, constrained by local laws, or when a species is widespread to the extent that comprehensive sampling is extremely complex and costly (Raxworthy and Smith 2021; Roos et al. 2021). Museum collections, therefore, represent an unexplored treasure trove preserving easily accessible genomic variability.

2.1.3 – East African Baboons

Museomics emerges as a valuable tool for investigating the evolutionary history of baboons (genus *Papio*), given the outlined considerations (Roos et al. 2021). Particularly significant are the olive and hamadryas species, which share overlapping ranges in East Africa. Specifically, *P. anubis* (FIGURE 2.2A) has been observed from Western to Eastern Africa, spanning the sub-Saharan region up to the Great Lakes area, excluding tropical forest zones (Jolly 1993). Conversely, *P. hamadryas* (FIGURE 2.2B) has been observed in the Horn of Africa and the South-Western coast of the Arabian Peninsula, distinguishing itself as the sole baboon species found beyond the confines of the African continent (Jolly 1993).

A



B



FIGURE 2.2 - Eastern African male baboons. A) *P. anubis*, photo by Roger Smith. Image via Flickr licensed under Creative Commons CC BY-NC 2.0 (<https://www.flickr.com/photos/wodjamiff/>). B) *P. hamadryas*, photo by Brian Henderson. Image via Flickr licensed under Creative Commons CC BY-NC 2.0 (<https://www.flickr.com/photos/stinkenroboter/>)

The contact zone between the two species is in Ethiopia, along the Awash River in the Awash National Park. In this area, ongoing hybridisation events between the two species have been observed (Nagel 1973; Newman 1997; Bergman, Phillips-Conroy, and Jolly 2008). This is very interesting due to differences in social organisation and male mate behaviour. Hamadryas baboons typically engage in mating within harems or one-male units (OMUs), with males

exhibiting sustained interest in adult females regardless of their reproductive state. Male hamadryas establish and maintain long-term bonds with females, maintaining the cohesion of their OMUs. The leader male of the OMU has exclusive sexual access to females of his group. The dispersion is typically on the female side, while males are philopatric. Conversely, olive baboons live in multi-male, multi-female groups, with male-female interactions limited to the receptive period of females. Male olive baboons primarily compete for females in oestrous, forming short-term consortships during which they vigorously maintain proximity to oestrous females. On the other hand, however, females can mate with different males during their ovulation period. Females usually never leave their natal group forming a matrilinear stable hierarchy, while males disperse from their group (Bergman, Phillips-Conroy, and Jolly 2008).

Hybridisation in the Awash region is driven by the cross-migration of males from both species into mixed or heterospecific groups, as well as by group fusion (Sugawara, 1979). Numerous hybrid and backcrossed individuals have been observed in the region (Phillips-Conroy & Jolly, 1986). Despite their presence, there appears to be some form of selection against hybrids (Nagel 1973; Nystrom 1992; Beyene 1998; Bergman 2000; Beehner 2003), as the hybrid zone has remained narrow, contrary to what one would expect given the known dispersal distances of baboons. Additionally, there is a wide variety of hybrid forms, indicating that hybrids do not have a dominant form, as no single hybrid type is favoured. Besides lacking an evident genetic advantage, hybrids do not possess behavioural advantages either, as they exhibit intermediate characteristics of the two species. For instance, hybrid males attempt to form OMUs with behaviours similar to pure hamadryas males, but they fail to maintain control and cohesion of females within their group. Consequently, other males mate with their females when they come into estrus (Sugawara 1988). Similarly, hybrid females tend to form OMUs, but without the strict adherence seen in hamadryas females, reducing their reproductive success (Beehner and Bergman 2006).

Furthermore, it has been observed that female behaviour is much more flexible than that of males. Hamadryas males in olive baboon groups tend to form stable bonds with any seemingly unattached females, while olive baboon females in hamadryas groups quickly learn to follow the OMU leader (Kummer et al. 1970; Nystrom 1992). Consequently, the reproductive success of males from both species is most favourable within their respective social contexts, and thus, sexual selection acts against mixed mating strategies, keeping the hybrid zone under control (Nystrom 1992).

Mitochondrial DNA analysis of the two species has shown that there may be an introgression of hamadryas' mitochondrial DNA in olive baboons. Roos et al (2019) observed the presence of Ethiopian olive baboons whose mitochondrial DNA forms a sister haplogroup to a hamadryas haplogroup from Somalia, indicating the possibility of ancient hybridisation events (FIGURE 0.4). Furthermore, mitochondrial genome analysis revealed that there is a structure among olive populations that reflects their geographical distribution. In particular, olive mitochondrial haplotypes from East, Central and West are separated as well and hamadryas haplotypes from different areas of Eastern Africa are each in a different branch (FIGURE 0.4; Zinner et al. 2013; Roos et al. 2021). Furthermore, the Ethiopian haplotype of the olive baboon group with the mtDNAs present in hamadryas, particularly in the clade with Ethiopian hamadryas and yellow Tanzanian baboons (Zinner et al. 2013; Roos et al. 2021). Mitochondrial haplotypes of Central African olive baboons form a sister clade to the one previously described, and this macro-clade

is sister to the clade composed of haplotypes of Western African olive and Guinea baboons (FIGURE 0.4; Zinner et al. 2013; Roos et al. 2021).

No in-depth studies are focusing on the nuclear DNA population structure of these two species. Concerning *P. hamadryas*, several nuclear genomes are available, but all have been extracted from individuals originating from a single location, Filhoa in the Awash National Park (Chiou et al. 2022; Sørensen et al. 2023). For *P. anubis* baboons, while many genomes are available, no study has yet explored their population structure properly. This is likely because these genomes have been derived only from individuals originating from a few localities between Tanzania and Kenya or living in captivity (Wall et al. 2016; Robinson et al. 2019; Rogers et al. 2019; Sørensen et al. 2023). The sole exception is the study by Sørensen (2023), who published several genomes from olive baboons inhabiting the Gog region in Western Ethiopia. In this particular study, it was observed that there is indeed genetic structure between the different populations of *P. anubis* analysed. For instance, in their ADMIXTURE analysis, the Gog individuals displayed a unique ancestry compared to olive baboons from Tanzania and Kenya who exhibited varying proportions of other ancestries, with a significant contribution from yellow baboons from the Tarangire population (Tanzania). In contrast, populations from Gombe, Serengeti (Tanzania), and Aberdare (Kenya) also showed ancestry from Ethiopian olive baboons, in addition to the Southern olive ancestry that was exclusively found in the Lake Manyara and Ngorongoro populations (Tanzania).

Thus, Sørensen's findings suggest that the olive baboon populations between Tanzania and Kenya have undergone differential hybridisation with yellow baboons. According to their results, the Tarangire population has had, and/or is currently experiencing, gene flow with the Eastern Tanzanian yellow baboons (excluding the Mikumi population, which appears to lack olive ancestry). This Eastern population also has interactions with the Western Tanzanian yellow baboons, which in turn show traces of gene flow with kinda baboons and the Western Tanzanian olive baboons of Gombe (Tanzania). These Gombe baboons, in turn, interact with the Northern Tanzanian olive populations near Tarangire, such as those in Lake Manyara. Additionally, it is notable that all of these populations form the same clade from a mitochondrial perspective, without species distinction, except for the Western Tanzanian yellow baboons, which group with kinda and chacma baboons (Sørensen et al. 2023).

Just a few kilometres to the North, across the border between Tanzania and Kenya, lies a famous and well-studied hybridisation site, the Amboseli National Park, between olive and yellow baboons (in the form of the Ibean yellow baboons, which inhabit this geographic area). These Ibean yellow baboons exhibit relatively coarse and wavy olive-like mane hair compared to typical yellow baboons (Jolly 1993).

The first evidence of hybridisation in the Amboseli region, between Simba Springs and Ithumba Hill, was documented in the 1960s (Maples & McKern, 1967). Animals in the region exhibited morphological traits associated with both olive and yellow baboons (Maples, 1972), with animals near Ithumba Hill showing mostly olive features and those near Simba Springs displaying mostly yellow traits. Despite the area's proximity to the boundary between the distributions of olive and yellow baboons, no olive baboons or olive-yellow hybrids were detected until the early 1980s. At that time, an unhabituated group comprising olive, yellow, and phenotypically mixed juveniles moved into the area (Samuels & Altmann, 1986). In a genetic study of

introgression in the greater Amboseli area, Charpentier et al. (2012) found that, consistent with phenotypic observations, the hybrid zone is relatively narrow, with asymmetric gene flow from olive into yellow populations. This suggests a reproductive advantage biased in favour of olive. Hybrid olive-yellow baboons appear to differ from yellow baboons in certain life history traits, such as an earlier age of male dispersal/migration (Alberts & Altmann, 2001; Charpentier et al., 2008) and an earlier age of maturation (Charpentier et al., 2008). Both of these characteristics may contribute to the reproductive advantage observed in the hybrid zone. However, based on longitudinal genetic data, Tung et al. (2008) found that the proportion of olive genetic background in hybrids decreased over time rather than increased. They, therefore, concluded that the reproductive and life history advantages associated with the olive background must interact dynamically with non-selective processes, such as stochastic variations in the rate of olive immigration into Amboseli.

Since hybridisation between olive and yellow baboons has already been well-characterised from a genetic perspective, this chapter will only address it as confirmation to validate the analyses, without attempting to uncover anything novel. Instead, the focus of this chapter will be on the hybridisation between hamadryas and olive baboons, and the potential effects of this hybridisation combined with the Southern hybridisation with yellow baboons on the structure of olive and potentially also hamadryas baboons. To this purpose, a comprehensive dataset has been assembled, consisting of modern and museum genomes, many of which are unpublished, of hamadryas and olive baboons. This dataset aims to cover the widest possible range of both species, with particular attention given to the well-known hybridisation zone of central Ethiopia.

2.2 - Material and Methods

2.2.1 - DNA Collection, Extraction, Sequencing and Final Dataset

We collected fourteen teeth from an equal number of skulls hosted by two Italian museums: “Giacomo Doria” Civic Museum of Natural History of Genoa and the Milan Natural History Museum. Of these, five were assigned to species *P. anubis* (FIGURE 2.3A), eight to species *P.*

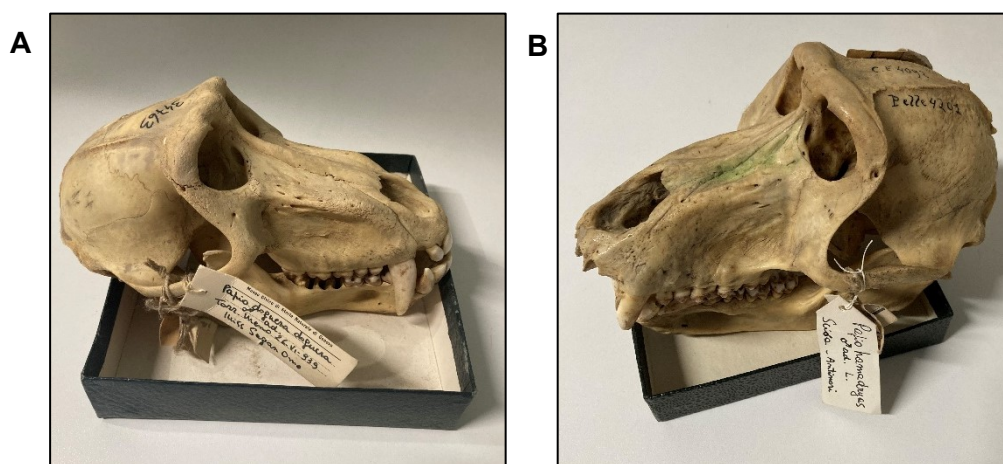


FIGURE 2.3 - *Papio* skulls. A) *Papio anubis* (male; MSNG34763) and B) *Papio hamadryas* (male; MSNG4092) skulls. From Natural History Museum “Giacomo Doria” in Genoa. Photo by Matteo Caldon.

hamadryas (FIGURE 2.3B), and one to species *P. cynocephalus*; of these, four have an unknown locality of origin (FIGURE 2.4; Supplementary Table 2). The specimens were collected between XIX and XX centuries in various expeditions as previously mentioned (Supplementary Table 2). DNA extraction and sequencing were carried out by colleagues at the Institute of Genomics, University of Tartu, Estonia, following the procedure outlined in Saupe et al. (2021).

In addition to the newly generated genomic museum data, we incorporated genomes from both museum and modern specimens recovered from literature. Specifically, we included museum specimens from Roos et al. (2019) and Grathwol et al. (2023), which feature ten hamadryas and ten olive baboons from various locations within the ranges. For modern samples, we added seven olive baboons from Robinson et al. (2019), which comprise a selection of founders from the SNPRC colony captured in Southern Kenya (including outliers 1X0812 and 1x4384), and nine wild baboons from two locations in Kenya (Rogers et al. 2019; Wall et al. 2016). Sample 1x4384 in Robinson's (2019) original study is recorded as *P. anubis*, but this is probably a labelling error. In fact, it is listed as *P. cynocephalus* in the genomic databases. Additionally, we included twenty-four olive baboons from six localities in Tanzania (four per locality; Sørensen et al. 2023), 4 olive baboons from Gog, Ethiopia (Sørensen et al. 2023), and 8 hamadryas baboons from Filhoa, Ethiopia (Sørensen et al. 2023; Chiou et al. 2022; FIGURE 2.4; Supplementary Table 2). Finally, we incorporated a single *P. cynocephalus* sample from Mikumi, Tanzania (Sørensen et al. 2023), the least admixed population that has been identified, and a *Man-drillus leucophaeus* (drill) genome as an outgroup (from PRJNA785018 BioProject in NCBI; Supplementary Table 2).

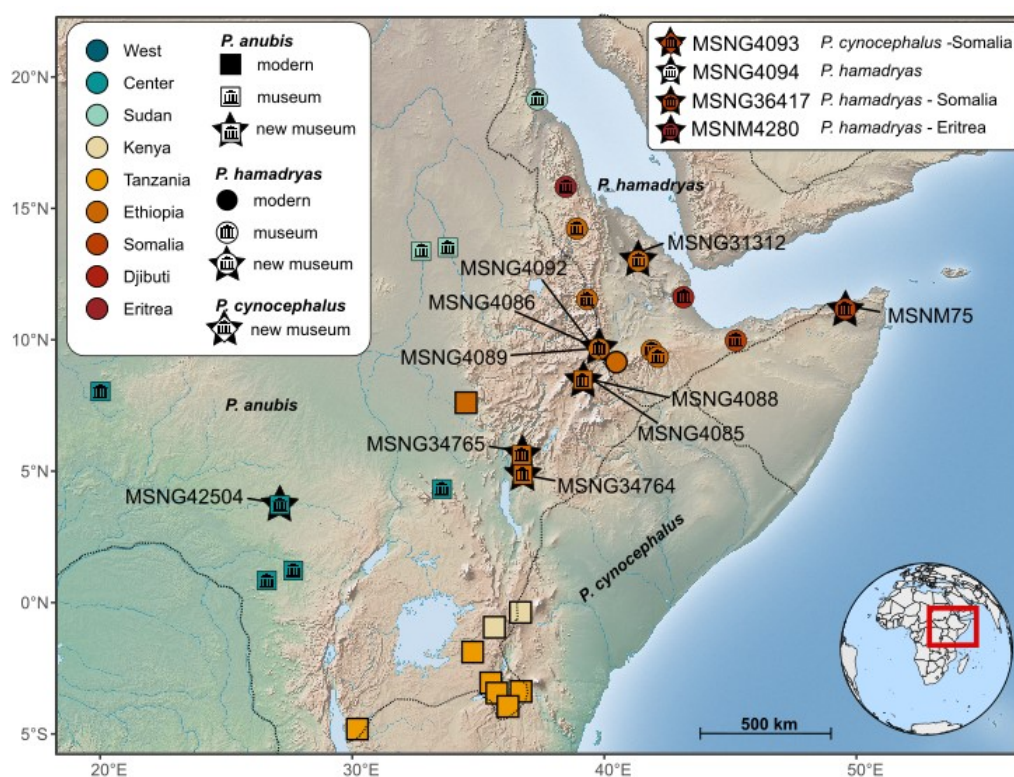


FIGURE 2.4 - Map with the locations of origin of modern populations and museum specimens of baboons. The newly generated museum specimens are highlighted with the museum code and a star. More details are in Supplementary Table 2.

2.2.2 - Genome Mapping

All the reads were mapped with BWA-MEM2 v.2.2.1 (Vasimuddin et al. 2019) to the *Macaca mulatta* reference genome (Mmul_10;). We then marked the duplicate reads with Picard Mark-Duplicates version 2.8.1 (cit.) and, only for medium-high coverage genomes, filtered the result using the option -q 10, -F 1292 and -f 2 of samtools view (Danecek et al. 2021).

2.2.3 - Variant Calling

We called variants for the modern specimens through the software GATK version 4.2.4.1 (McKenna et al. 2010). We generated the gVCF using HaplotypeCaller and performed joint calls via GenotypeGVCFs. The resulting VCF was then filtered using VariantFiltration, applying stringent filters for SNPs ("QD < 2.0 || MQ < 40.0 || FS > 60.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"). Using PLINK v.1.9 (Chang et al. 2015), all sites from the sex chromosomes and with a minor allele frequency (MAF) below 0.05 were removed, keeping only the biallelic SNPs filtered using the --snps-only option. The final set is composed of 24,943,534 autosomal SNPs. For all the other samples (the museum and the low-coverage samples, see Supplementary Table 2) we called the variants through with *pileupCaller* (Lamnidis et al. 2018; <https://github.com/stschiff/sequenceTools>), using the panel of SNPs previously generated.

2.2.4 - Autosome Analysis

PCA

We performed a preliminary survey of population structure in the set of samples by running a principal component analysis (PCA) using the smartpca function implemented in EIGENSOFT software 8.0.0 (Patterson, Price, and Reich 2006). We performed these PCAs for the high-coverage samples; the low-coverage and museum specimens were instead projected using lsqproject set to YES. All the other settings are used with default parameters.

ADMIXTURE

To understand the genomic components present in individuals belonging to the three North-Eastern *Papio* species, we perform ADMIXTURE analysis using the software ADMIXTURE v.1.3.0 (Alexander, Novembre, and Lange 2009), applying the --cv option to estimate the cross-validation values. We run the analysis for K between 2 and 6. We also performed analyses applying the --supervised option to avoid potential biases that may arise due to the high number of individuals with very low coverage, indicating which individuals will lead the analysis and how many different ancestries to use. We performed supervised ADMIXTURE with K from 2 to 4. To decide which samples to use to lead the analysis, we based our decision on Sørensen's (2023) results.

K-means clustering and D-Statistics

Samples were grouped using an unsupervised clustering approach. Cluster analysis, or clustering, is an unsupervised classification technique where a dataset - often multidimensional - is grouped into clusters so that items within each cluster are similar to each other according to a specific criterion. K-means, a widely-used iterative procedure, partitions N objects into K distinct clusters. It is perhaps the most popular clustering method and is particularly well-known among partition-based clustering techniques, which use centroids to represent clusters (Oti et al. 2021).

Here the clustering was performed using the results obtained from the smartpca analysis, with the k-means command from the *cluster* package (Maechler 2019) in R version 4.3.2 (R Core Team 2021). The command was used with `nstart = 100` to increase the number of initialisations, and using `K=6` as this was the optimal result according to the Silhouette index (Shutaywi and Kachouie 2021). This index measures how similar each data point is to the points within its own cluster compared to points in other clusters. It ranges from -1 to 1, where higher values indicate more cohesive clusters.

After the assignment of individuals to each cluster, D-statistics were performed first to better evaluate the structure of *P. anubis* by considering clusters of olive as H3, H2, and H1, and then to assess the presence of signals of differential gene flow between olive (H2 and H1) and the two neighbouring species, yellow and hamadryas baboons (H3).

D-statistics were performed using qpDstat (Patterson et al. 2012), with default parameters, to test for the occurrence of any imbalance in allele sharing between different populations and species (H1, H2 and H3), including drill (*M. leucophaeus*) as an outgroup (H4). The observations are considered significant if $|Z|$ -score > 3.

f4-ratio test

To infer the mixing proportions of an admixture event between the olive baboon clusters and *P. hamadryas*, we calculated the f4 ratio through qpF4ratio implemented in AdmixTools (Patterson et al. 2012) with the form $\alpha = f4(A,O; X,C) / f4(A,O; B, C)$, including drill (*M. leucophaeus*) as an outgroup (O).

2.3 - Results

2.3.1 - Population Structure of Eastern African Baboons

We initially explored the dataset by using principal component analysis (PCA): principal component 1 (which explains 16.89% of the amount of variation) drives the separation between hamadryas and olive/yellow baboons, while the partition between the hamadryas/olive and the yellow baboons is present along the principal component 2 (9%) (FIGURE 2.5A). In this plot, we observe that the yellow baboon samples are well separated from the hamadryas and olive baboon samples, which appear to be much closer to each other. Additionally, we note that the olive baboon samples from the Tanzanian population of Tarangire seem to be more attracted by *P. cynocephalus* samples. Instead, the museum *P. cynocephalus*-labelled sample from Somalia (MSNG4093) doesn't plot with yellow baboon samples but with the hamadryas baboon samples, suggesting it could have been incorrectly assigned to that species.

A second PCA was generated using only hamadryas and olive baboon samples (FIGURE 2.5B). In this case, principal component 1 (20.68%) drives the separation between hamadryas and olive samples, while the olive baboon populations are spread out along principal component 2 (5.94%), suggesting some degree of population structure. Specifically, we observed at least three groups that resemble in part the geographic origin of the samples. The first group at the top of the plot, is composed of most of the populations from Tanzania (Arusha, Lake Manyara,

Ngorongoro and Tarangire which occupies the extreme end of the plot). The second group is plotted vertically on the opposite side of the plot and is composed of the olive baboons from Ethiopia, both museum and modern specimens, as well as a museum specimen from Sudan (SAMEA112757473). Among them plot also 1X0812 and two museum specimens assigned to *P. hamadryas* (MSNM4280 and SAMEA112757480). A third group of individuals is located just above the Ethiopian olive baboons, consisting of other museum specimens from Sudan, Congo, and Western regions, as well as the individuals from Gombe (Tanzania). Closely associated with this group but slightly leaning towards the cluster of Tanzanian samples are the individuals from Aberdare and Masai Mara (Kenya), the founders of the SNPRC colony (which were captured in Kenya), and the individuals from Serengeti (Tanzania). We note also that no major subdivision among *P. hamadryas* samples is detected. All samples were plotted very close to each other forming a cloud. As in the previous PCA, the sample MSNG4093 from Somalia, which was assigned to *P. cynocephalus*, clusters close to the hamadryas group, although it remains somewhat separated.

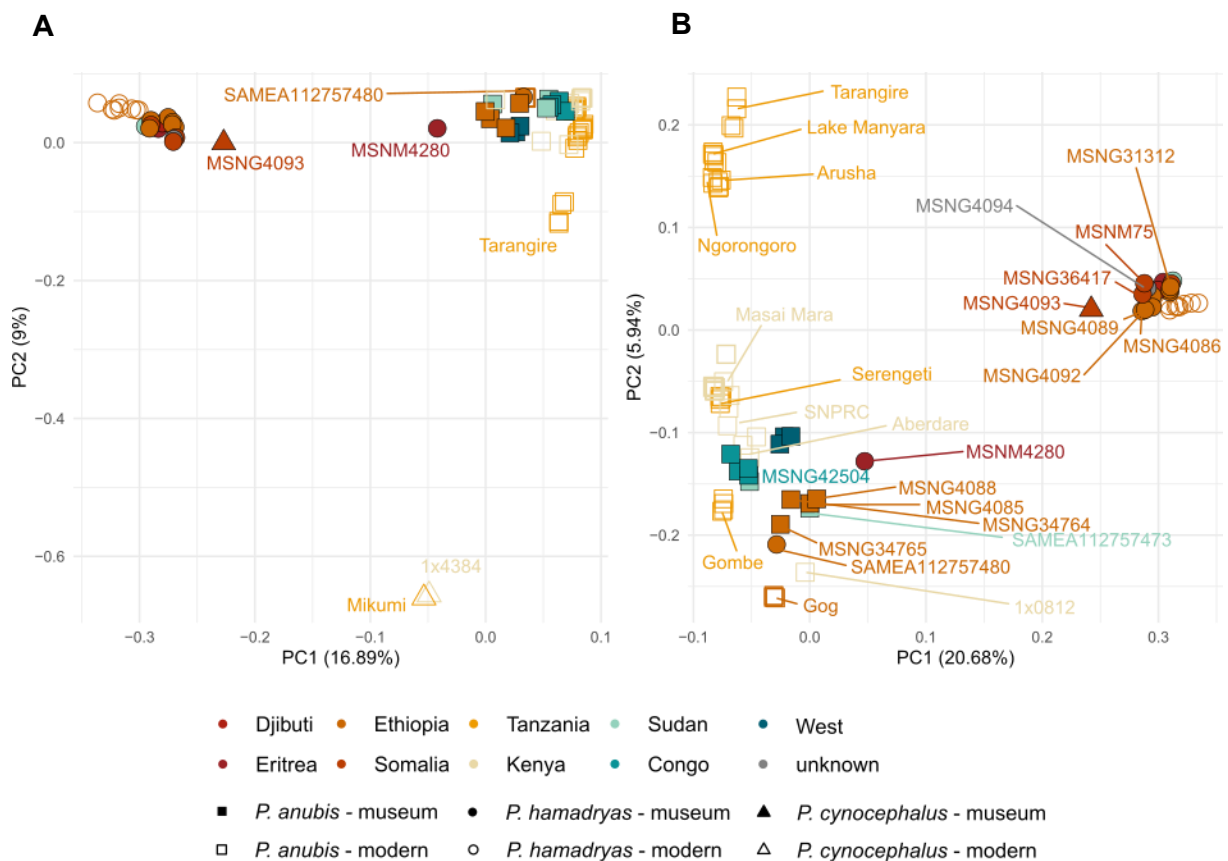


FIGURE 2.5 - PCA analyses on Eastern Africa samples. A) Analysis using *P. anubis*, *P. hamadryas* and *P. cynocephalus* samples: due to their association with the species, the museum specimens found in unexpected positions in the PCA have been highlighted; the yellow SNPRC and Mikumi specimens and the olive population from Tarangire are also highlighted. B) Analysis using *P. anubis* and *P. hamadryas* samples only: the newly generated museum specimens, those with particular positions in the PCA, and the modern populations have been highlighted. In A and B shapes refer to species; empty shapes refer to modern specimens, and filled shapes to museum specimens; the colours refer to the region of origin (true or presumed) of the specimens.

ADMIXTURE analysis was initially conducted using the complete dataset of only olive and hamadryas baboons, excluding yellow baboons; K values from 2 to 6 were tested. In this, the most supported K value was 2 (CV = 0.64; FIGURE 2.6A). For K=2 separation between hamadryas and olive baboon can be observed. In most cases, individuals exhibited only one genomic component, though some individuals showed a second minor component too. The olive baboons showing a small hamadryas component (HC) were the ones from Ethiopia (both museum and modern specimens), the Robinson outlier 1X0812, one Sudanese sample (SAMEA112757473), olive baboons from West Africa (aDNA97, aDNA95, aDNA94), and - though with a tiny contribution - modern samples from Tarangire in Tanzania and Aberdare in Kenya. As for the hamadryas baboons showing the olive component (OC), only two museum specimens, exhibited this pattern: SAMEA112757480 showed an OC with a small hamadryas contribution (comparable to Ethiopian olive samples), while MSNM4280 displayed a profile with almost 50:50 in the OC and HC. These two samples were the same that in the PCA (FIGURE 2.5) clustered with olive baboon samples.

Using the same dataset, a supervised ADMIXTURE analysis was subsequently conducted, in order to prevent biases arising from the low coverage of many genomes (FIGURE 2.6B). The individuals chosen to lead the analysis for olive baboons were individuals from Lake Manyara and Ngorongoro, who were 100% OC in both the previous result and in Sørensen's (2023) ADMIXTURE analysis. For hamadryas, four individuals from Filhoa were used, as they too were 100% HC in the previous results and in Sørensen's (2023) analysis. The outcome of this supervised analysis was largely identical in trend to the previous one, showing most samples having a single component and few a minor second component.

The next step involved testing the dataset via ADMIXTURE analysis, including two yellow baboon samples. Again, K=2 was the most supported value (CV = 0.41; FIGURE 2.6C). The trend remained the same, with the same individuals showing mixed components. In this case, the two yellow baboon samples appeared to consist of half OC and half HC. We then performed supervised ADMIXTURE analysis (K=3; FIGURE 2.6D) fixing the two yellow baboons as a main source (YC), along with the same reference groups as before - olive baboons from Lake Manyara and Ngorongoro, and hamadryas baboons from Filhoa. This analysis showed some differences compared to the previous ones. All Ethiopian olive baboons (including 1X0812) and the Sudanese baboon displayed about 20% HC, with the remaining part being OC. The Western olive baboons retained about 10% HC, with the rest being OC and a very small percentage of YC. The olive baboons from Tarangire showed an increased percentage of non-OC, around 25%, but this time it was entirely YC. Olive baboons from Arusha also appeared to have about 10% YC, as did some individuals from Masai Mara and Aberdare. Gome olive baboons exhibited a minimal percentage of YC. Hamadryas baboons were consistently 100% HC, except for the two museum specimens SAMEA112757480 and MSNM4280, which displayed percentages similar to the previous results, with a predominance of OC over HC.

In Sørensen's (2023) analysis, Ethiopian olive baboons exhibited a distinct ancestry from the others (see Figure 2D in Sørensen et al. 2023), and so further supervised ADMIXTURE analysis was conducted using K=4 (FIGURE 2.6E). The groups used as a reference for yellow, hamadryas and olive components remained the same as before, but a fourth leading group was added - four olive individuals from Gog (EOC), as in Sørensen's analysis (2023). The result showed that, with regards to the hamadryas baboons, no changes were observed except for the two

museum specimens: SAMEA112757480 was composed by a unique component (EOC), while MSNM4280 became a combination of two component where EOC predominated over HC. The Ethiopian olive baboons were composed of 100% EOC, except for two museum specimens, MSNG4085 and MSNG4088, which also displayed about 10% HC. All central olive baboons (from Congo, etc.), most Masai Mara baboons, and West and Sudan baboons (except for aDNA94 and SAMEA112757473, who showed a small percentage of HC) were also 100% EOC. The Robinson outlier (1X0812) was a combination of EOC (predominantly), about 15% HC, and 5% OC. The other founder individuals from the SNPRC colony were composed of roughly equal parts OC and EOC, as were the Gombe and Serengeti baboons. Aberdare individuals predominantly showed EOC (around 60%), compared to OC, but one of the two also had about 10% YC. Arusha individuals were 100% OC, except for one which had 5% YC, while Tarangire individuals were OC with approximately 20% YC.

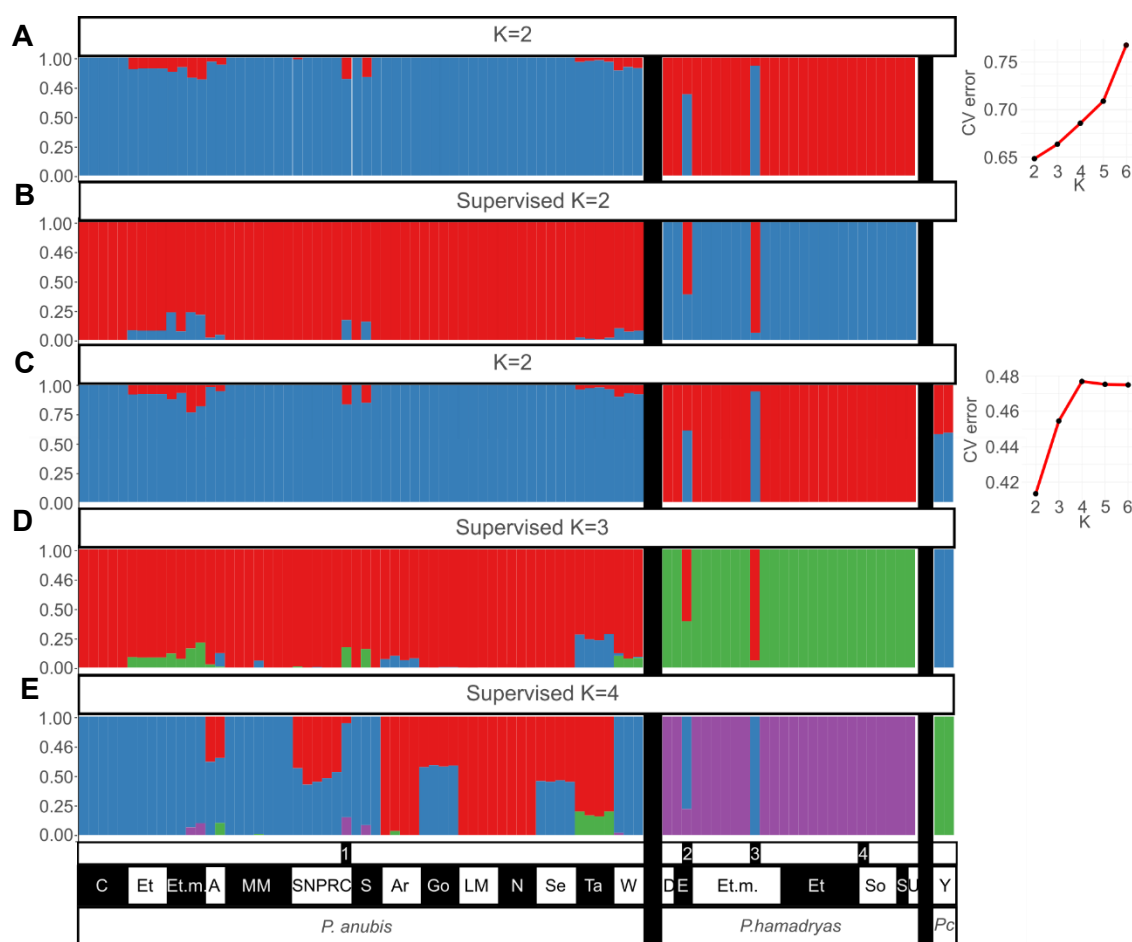


FIGURE 2.6 - ADMIXTURE analyses of Eastern African baboons. A) Unsupervised analysis with the most supported K values (K=2) and the corresponding cross-validation (CV) curve for the different K values (on the right); B) supervised analysis with K=2. In A and B analysis was performed using only olive and hamadryas baboon samples. C) Unsupervised analysis with the most supported K values (K=3) and the corresponding cross-validation (CV) curve for the different K values (on the right); D) supervised analysis with K=3 and E) with K=4. In C, D and E analysis was performed using yellow, olive and hamadryas baboon samples; in B, C and D the populations used as leading references for the analysis are for K=2 LM/N and Et (*P. hamadryas*); for K=3 LM/N, Et (*P. hamadryas*) and Y; for K=4 Et (*P. anubis*), LM/N, Et (*P. hamadryas*) and Y. The sample origin is reported on the bottom; the numbers below the plots are referred to as 1= 1X0812; 2= MSNM4280; 3= SAMEA112757480; 4= MSNG4093; the abbreviations refer to C= Congo; Et= Ethiopia modern; Et.m= Ethiopia museum; A= Aberdare; MM= Masai Mara; S= Sudan; Ar= Arusha; Go= Gombe; LM= Lake Manyara; N= Ngorongoro; Se= Serengeti; Ta= Tarangire; W= western; D= Djibouti; E= Eritrea; So= Somalia; U= unknown origin; Y= yellow baboon; Pc= *Papio cynocephalus*.

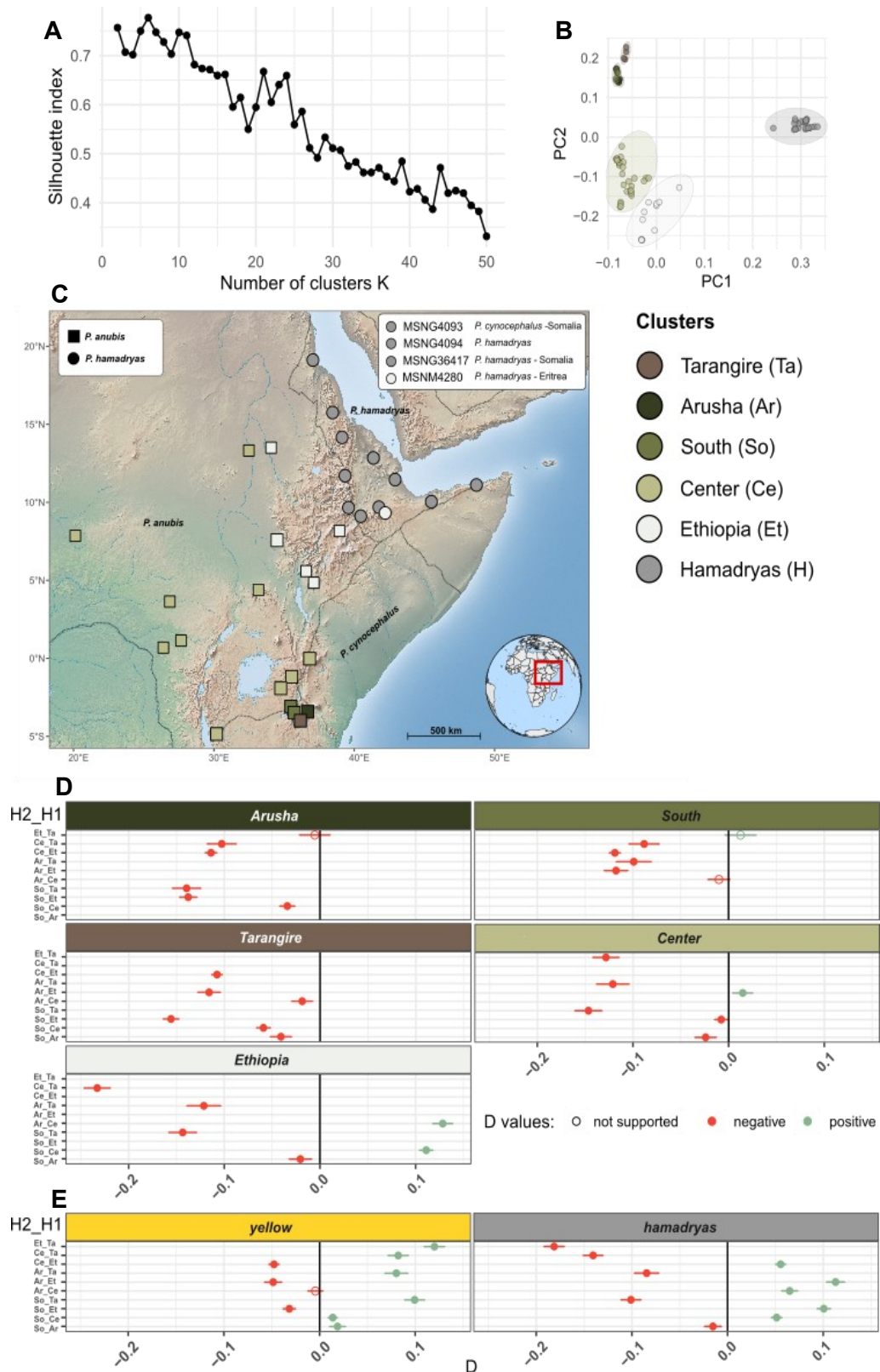


FIGURE 2.7 – K-mean clustering and D-statistics. A) Silhouette index variation for the number of clusters; B) PCA showing the clusters division, the colours refer to the legend below; C) map with locations and the associated cluster (the colours refer to the legend on the right); D) D-statistics using the *P. anubis* clusters as H3, H2, and H1; E) D-statistics using yellow and hamadryas samples as H3 and *P. anubis* clusters as H2 and H1. In C and D, the abbreviations refer to the legend above. The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. The population name at the top of each panel refers to the yellow population used as H3. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2.

2.3.2 - Hybridisation Events

Following these results, a formal clustering of olive and hamadryas individuals was attempted in order to rigorously classify the individuals. Clustering was performed using K-means, and the Silhouette index was tested for this data, showing that the best K value was six (FIGURE 2.7A). K-means clustering results for K=6 indicated that hamadryas formed a distinct cluster (H), with the exception of the two museum samples, SAMEA112757480 and MSNM4280, which were placed in a second cluster composed of Ethiopian olive baboons, SAMEA112757473 from Sudan, and 1X0812 from SNPRC (Et). Another cluster included all the other SNPRC individuals, all samples from Kenya, Central and West Africa, and individuals from Serengeti and Gombe (Ce). A fourth cluster consisted of individuals from Lake Manyara and Ngorongoro (So), a fifth comprised Arusha individuals (Ar), and the sixth cluster included Tarangire individuals (Ta; FIGURE 2.7BC; Supplementary Table 2).

These clusters were used as populations to be tested using D-statistics in order to have a formal subdivision of individuals, rather than basing grouping on country of origin individuals. The first analysis involved D-statistics among the olive clusters (FIGURE 2.7 D). The results suggested that cluster So shared more alleles with the Ce and Ar clusters than with the Ta and the Et clusters; in turn, cluster Ce shared more alleles with cluster So, followed by the Et, and less with Ar and Ta. Cluster So shared more alleles with cluster Ce; Ta and Et showed no significant difference in allele sharing with this cluster. Cluster Et, in turn, showed greater allele sharing with cluster Ce, and less with Ta. Cluster Ta shared more alleles with So, and less with Et.

Subsequently, hamadryas was tested as H3, with the olive clusters assigned as H1 and H2. The results suggest that hamadryas shares more alleles with the Et cluster than with all the others; the least allele sharing is with Ta cluster. On the other hand, when yellow is used as H3, it is observed, as also noted by Sørensen (2023), that the population with the strongest signal of gene flow is Tarangire. Meanwhile, the clusters showing the least allelic sharing are the Et and So olive clusters (FIGURE 2.7E).

The relative proportion of components in olive baboons was characterised using the f4-ratio (FIGURE 2.8). The position of populations in the phylogenetic tree was determined based on the results of the ADMIXTURE and D-statistics, as well as observations from Sørensen (2023). Therefore, cluster Ta was used as external population (A), with the So and Ar clusters alternately as population B. Hamadryas was consistently used as population C, and the Et and Ce clusters were used as test population X, together with one of the others three olive clusters that were not alternatively used as population A or B (FIGURE 2.8). The results showed that in all cases, the Et cluster exhibited the lowest proportion of olive component relative to hamadryas presence, confirming the observations from the D-statistics. The Et cluster had values around 70-75% olive component, compared to over 90% observed in the other tested clusters. The only case where the tree structure was unsupported was when the So cluster was used as X and Ar as B, suggesting that Arusha has a lower proportion of olive component than the So cluster.

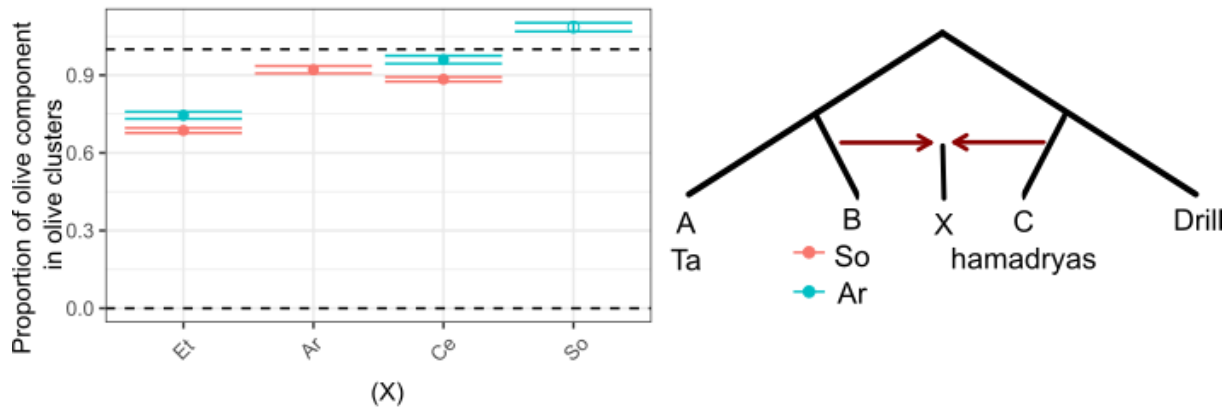


FIGURE 2.8 - Proportion of olive component in olive baboon clusters. F4-ratio analysis testing the relative presence of *Ta* component in the other clusters. The colours refer to the cluster used as the B branch (as described on the right). The tree topology is indicated on the right. The empty dots refer to alpha values associated with unsupported Z values; the bars show the extension of three standard deviations. Abbreviations refer to clusters Ethiopia (Et), Centre (Ce), South (So), Arusha (Ar) and Tarangire (Ta).

2.4 - Discussion

Our findings are consistent in suggesting that there is no evident population structure in *Papio* hamadryas. These results align with previous studies that employed mitochondrial DNA fragments, which similarly suggested a lack of structured differentiation. Specifically, individuals from the same geographic regions were found to cluster inconsistently, resulting in clades that included individuals from distant areas across East Africa (Kopp et al. 2014a). This inconsistent clustering was observed not only in partial mitochondrial sequences but also in full mitochondrial genomes derived from museum specimens (Grathwol et al. 2023). Consequently, it seems that African hamadryas baboon populations do not form distinct, monophyletic geographic clusters. This phenomenon can likely be attributed to the species' female-biased dispersal patterns. Females in *P. hamadryas* tend to migrate between groups more frequently than males, which disrupts the geographic correlation with mitochondrial genetic structure. As a result, genetic exchange among females diminishes the geographic boundaries that might otherwise create distinct mitochondrial lineages (Kopp et al. 2014a; Kopp et al. 2014b). As with the mitochondrial DNA, the absence of structure in the nuclear genome can also be attributed to the interconnectedness of the groups, which prevents the clear emergence of population structure.

Population structure in olive baboons is evident. Previous mitochondrial DNA analysis revealed clear genetic differentiation, which is to be expected given the species' extensive geographical range. Those studies have identified at least five distinct haplogroups (Roos et al. 2021). One of these haplogroups is shared with *P. cynocephalus* (yellow baboons) and *P. hamadryas*, and another with *P. papio*, while the remaining haplogroups are unique to *P. anubis* populations in Central and West Africa. Given that *P. anubis* dispersal is male-biased, it is unsurprising that mitochondrial DNA, maternally inherited, reveals a well-defined population structure.

This mitochondrial structuring is further supported by nuclear genome data, which shows that Ethiopian *P. anubis* populations form a distinct cluster, separate from other *P. anubis* populations. Furthermore, the population structure within Tanzanian olive baboons is particularly

pronounced. For example, populations in Tarangire display significant hybridisation with yellow baboons, which distinguishes them from other Tanzanian olive populations. Indeed, the ancestry patterns suggest that *P. cynocephalus* has contributed genetically to some olive populations, particularly in regions such as Tarangire and Arusha as previously reported (Sørensen et al. 2023).

In contrast, Ethiopian olive baboons stand out as a distinct group in both mitochondrial and nuclear analyses (Roos et al. 2021; Grathwol et al. 2023; Sørensen et al. 2023). The results of this study suggest that this differentiation may be due to hybridisation with *P. hamadryas*. ADMIXTURE analyses provide robust evidence of *P. hamadryas* gene flow into these Ethiopian olive baboon populations, which is further corroborated by the use of D-statistics and f4 ratio tests that clearly indicate the presence of hamadryas ancestry in these groups.

Hybridisations between *P. hamadryas* and *P. anubis*, appear to play a key role in shaping the genetic diversity of these baboon populations and may have contributed to local adaptations. Behavioural barriers between *P. hamadryas* and *P. anubis*, while significant, appear insufficient to completely prevent hybridisation (Phillips-Conroy & Jolly, 1986). This suggests that ecological pressures, particularly in regions where their ranges overlap, may facilitate occasional genetic mixing.

Overall, the findings, in conjunction with existing research, suggest that the evolutionary history of African baboons has been profoundly shaped by episodes of hybridisation and introgression (FIGURE 2.9). This genetic admixture complicates the traditional phylogenetic view of baboon evolution, which often relies on morphological or behavioural traits to define species boundaries.

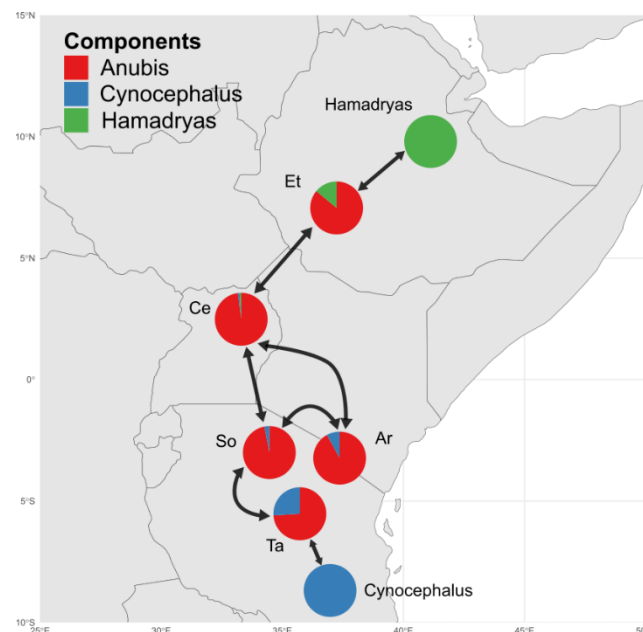


FIGURE 2.9 – Relationship in Eastern African baboons. Map combining ADMIXTURE and D-statistics results. Each pie chart represents a cluster (including yellow and hamadryas); the values in the pie chart are extracted from supervised ADMIXTURE analysis ($K=3$), by calculating the average proportion of each component for each cluster. The arrows refer to the D-statistics results: for each cluster as H3 was selected the cluster as H1 or H2 with which there is the highest signal of gene flow; the resulting couple was connected by an arrow. Abbreviations refer to clusters Ethiopia (Et), Centre (Ce), South (So), Arusha (Ar) and Tarangire (Ta).

Additionally, in Ethiopia, there is evidence of further hybridisation events, not only between *P. hamadryas* and *P. anubis* but also involving a species from a different genus, *Theropithecus*. Emerging evidence suggests the possibility of hybridisation between *P. anubis* and *Theropithecus gelada* (geladas), a phenomenon that has been observed both in captivity and in the wild (Dunbar and Dunbar 1974; Jolly et al. 1997; Markarjan, Isakov, and Kondakov 1974). In the subsequent chapter, this potential inter-generic hybridisation will be examined in greater detail, providing further insight into the evolutionary interactions between these primate species.

The inclusion of data from museum specimens has provided a broader perspective, although it did not lead to the identification of new population groups. This is largely because the olive baboon specimens were mostly from Ethiopia, though they represented different regions. For *P. hamadryas*, despite the fact that specimens came from a variety of geographic locations, no clear population structure was observed.

This research underscores the significant role of hybridisation in shaping the population structure of olive baboons in East Africa, but it also draws attention to an often-overlooked issue concerning the classification and conservation of museum specimens. The genetic data gathered in this study reveal that at least one specimen, previously identified as *Papio cynocephalus*, may have been misclassified and could, in fact, be a *Papio hamadryas* individual. Likewise, two specimens labelled as *P. hamadryas* consistently grouped with olive baboons (*P. anubis*) in multiple genetic analyses. These specimens may represent hybrids, complicating their taxonomic classification. To address this issue, a more thorough morphometric analysis, particularly of the cranial features of these specimens, would be necessary to reassess their classification.

For the sample MSNM4280 from Eritrea, two hypotheses are possible to explain the results: hybridisation or classification error. If the hybridisation hypothesis were confirmed, it could indicate the presence of hybridisation events in the Northern region between Eritrea and Sudan. In fact, both the ADMIXTURE analysis and clustering revealed that, of the three Sudanese samples, two clusters with samples from Congo and Kenya, while one sample (SAMEA112757473), which geographically has a more easterly origin (FIGURE 2.7; Supplementary Table 2), clusters with the Ethiopian samples. In the ADMIXTURE analysis with $K=4$, this sample from Sudan shows not only the Ethiopian olive component but also a small proportion of hamadryas component, suggesting hybridisation in the history of this population.

This finding raises a broader concern about the accuracy of taxonomic identifications in museum collections. This is particularly relevant for specimens collected several decades ago when species definitions were less precise and rigorous than they are today and when different taxonomic labels were often applied to the same species. Historically, classification relied heavily on morphological traits, which can be ambiguous, especially in cases of hybridisation. As a result, museum labels may not always reflect current scientific standards, leading to potential misidentifications.

This situation poses a challenge for both researchers and conservationists, as incorrect taxonomic assignments can affect our understanding of species distribution, evolution, and conservation priorities. This study, therefore, highlights the importance of re-examining historical

museum specimens with modern techniques to ensure more accurate taxonomic classifications and to better inform conservation efforts.

2.5 - Conclusion

For the first time, this study provides an extensive investigation into the population structure of the two baboon species *P. anubis* and *P. hamadryas*. A substantial number of new genomes from various regions of East Africa have been added to the analysis, greatly enhancing our understanding of the genetic diversity and the underlying dynamics in play within these populations.

The study also thoroughly explores the impact of the well-documented hybridisation between the two species, examining how this interbreeding has influenced their population structure over time. By investigating these hybridisation dynamics, the research sheds light on the evolutionary processes shaping species boundaries and the complex genetic interactions between the two baboon species.

By incorporating a large dataset of new genomic information, this research offers a more comprehensive view of the genetic dynamics and provides critical insights into the evolutionary relationships between these baboon species. The findings highlight the importance of understanding hybridisation's role in shaping the evolutionary history and future of these species in East Africa.

2.6 - References

Alberts, S. C., & Altmann, J. (2012). The Amboseli Baboon Research Project: 40 years of continuity and change. In Kappeler, P.M., Watts, D.P. (Eds.), *Long-Term Field Studies of Primates* (pp. 261–287). Berlin: Springer-Verlag.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655–1664.

<https://doi.org/10.1101/gr.094052.109>

Ashby, J., and Machin, R. (2021). Legacies of Colonial Violence in Natural History Collections. *Journal of Natural Science Collections* 8:44–54.

Beehner, J. C. (2003). Female behavior and reproductive success in a hybrid baboon group (*Papio hamadryas hamadryas* x *Papio hamadryas anubis*) (PhD thesis). Washington University; Washington University

Beehner, J. C., & Bergman, T. J. (2006). Female behavioral strategies of hybrid baboons in the Awash National Park, Ethiopia. In Swedell, L., Leigh, S.R. (Eds.), *Reproduction and Fitness in*

Baboons: Behavioral, Ecological, and Life History Perspectives (pp. 53–79). New York: Springer.

Bergman, T. J. (2000). Mating behavior and reproductive success of hybrid male baboons (*Papio hamadryas hamadryas* X *Papio hamadryas anubis*) (PhD thesis). Washington University; Washington University.

Bergman, T. J., Phillips-Conroy, J. E., & Jolly, C. J. (2008). Behavioral variation and reproductive success of male baboons (*Papio anubis* x *Papio hamadryas*) in a hybrid social group. *American journal of primatology*, 70(2), 136–147. <https://doi.org/10.1002/ajp.20467>

Beyene, S. (1998). The role of female mating behavior in hybridization between anubis and hamadryas baboons in Awash, Ethiopia (PhD thesis). Washington University; Washington University.

Browne, J.E. (1992). A Science of Empire: British Biogeography before Darwin. *Revue d'histoire des sciences* 45, 453-475. <https://doi.org/10.3406/rhs.1992.4244>

Cerretti, C. 2003 “L'AFRICA E LA SOCIETÀ. GEOGRAFICA ITALIANA.” *Bollettino della Società Geografica Italiana* 12 (8): 423-431

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>

Charpentier, M. J. E., Tung, J., Altmann, J., & Alberts, S. C. (2008). Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology*, 17, 2026–2040.

Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., Barthès, N., Alberts, S. C., & Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioral observations in a hybrid population. *Molecular Ecology*, 21, 715–731.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>

Das, S. & Lowe, M. (2018). Nature Read in Black and White: decolonial approaches to interpreting natural history collections. *Journal of Natural Science Collections*, Volume 6, 4 - 14.

Doria, G., & Poggi, R. (2017). Annali del Museo Civico di Storia Naturale “Giacomo Doria”. *Erredi Grafiche Editoriali*. Genova

Dunbar, R. I. M., and Dunbar, P. (1974). On Hybridisation between *Theropithecus Gelada* and *Papio Anubis* in the Wild. *Journal of Human Evolution* 3 (3): 187–92.

Grathwol, F., Roos, C., Zinner, D., Hume, B., Porcier, S. M., Berthet, D., Cuisin, J., Merker, S., Ottoni, C., Van Neer, W., Dominy, N. J., & Kopp, G. H. (2023). Adulis and the transshipment of baboons during classical antiquity. *eLife*, 12, e87513. <https://doi.org/10.7554/eLife.87513>

Jolly, C. J. (1993). Species, Subspecies, and Baboon Systematics. In *Species, Species Concepts and Primate Evolution*, 67–108. *Springer US*, Boston.

Jolly, C. J., Woolley-Barker, T., Beyene, S., Disotell, T. R., & Phillips-Conroy, J. E. (1997). Intergeneric hybrid baboons. *International Journal of Primatology*, 18, 597–627.

Kopp, G. H., Ferreira da Silva, M. J., Fischer, J., Brito, J. C., Regnaut, S., Roos, C., & Zinner, D. (2014a). The Influence of Social Systems on Patterns of Mitochondrial DNA Variation in Baboons. *International journal of primatology*, 35(1), 210–225.
<https://doi.org/10.1007/s10764-013-9725-5>

Kopp, G. H., Roos, C., Butynski, T. M., Wildman, D. E., Alagaili, A. N., Groeneveld, L. F., & Zinner, D. (2014b). Out of Africa, but how and when? The case of hamadryas baboons (*Papio hamadryas*). *Journal of human evolution*, 76, 154–164.

Kummer, H., Gotz, W., & Angst, W. (1970). Cross-species modifications of social behavior in baboons. In Napier, J.R., Napier, P.H. (Eds.), *Old World Monkeys: Evolution, Systematics and Behavior* (pp. 351–363). New York: Academic Press

Lalueza-Fox C. (2022). Museomics. *Current biology: CB*, 32(21), R1214–R1215.
<https://doi.org/10.1016/j.cub.2022.09.019>

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khar-tanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., and Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature communications*, 9(1), 5018.
<https://doi.org/10.1038/s41467-018-07483-5>

Maechler, M. (2019). Finding groups in data: Cluster analysis extended Rousseeuw et al. R package version, 2(0), 242-248.

Maples, W. R., & McKern, T. W. (1967). A preliminary report on classification of the Kenya baboon. In Vagtborg, H. (Ed.), *The Baboon in Medical Research, Vol. 2* (pp. 13-22). Austin, TX: University of Texas Press.

Maples, W. R. (1972). Systematic reconsideration and a revision of the nomenclature of Kenya baboons. *American Journal of Physical Anthropology*, 36, 9-19.

Markarjan, D. S., Isakov, E. P., and Kondakov, G. I. (1974). Intergeneric Hybrids of the Lower (42-Chromosome) Monkey Species of the Sukhumi Monkey Colony. *Journal of Human Evolution* 3 (3): 247–55.

- Miriti, M. N., Rawson, A. J., and Mansfield, B. (2023). The History of Natural History and Race: Decolonizing Human Dimensions of Ecology. *Ecological Applications: A Publication of the Ecological Society of America* 33 (1): e2748.
- Nagel, U. (1973). A Comparison of Anubis Baboons, Hamadryas Baboons and Their Hybrids at a Species Border in Ethiopia. *Folia Primatologica; International Journal of Primatology* 19 (2): 104–65.
- Newman, T. K. (1997). Mitochondrial DNA analysis of intraspecific hybridization in *Papio hamadryas anubis*, *P. h. hamadryas* and their hybrids in the Awash National Park, Ethiopia (PhD thesis). New York University; New York University.
- Nystrom, P. D. A. (1992). Mating success of hamadryas, anubis and hybrid male baboons in a “mixed” social group in the Awash National Park, Ethiopia (PhD thesis). Washington University; Washington University
- Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive review of K-Means clustering algorithms. *criteria*, 12, 22-23.
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Phillips-Conroy, J. E., & Jolly, C. J. (1986). Changes in the structure of the baboon hybrid zone in the Awash National Park, Ethiopia. *American Journal of Physical Anthropology*, 71, 337–350.
- Podestà, G. L. (1998). Gli investimenti italiani in Africa orientale 1869-1919. In *Annali dell'Istituto storico italo-germanico in Trento*, 24, 147-189.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in ecology & evolution*, 36(11), 1049–1060. <https://doi.org/10.1016/j.tree.2021.07.009>
- Robinson, J. A., Belsare, S., Birnbaum, S., Newman, D. E., Chan, J., Glenn, J. P., Ferguson, B., Cox, L. A., & Wall, J. D. (2019). Analysis of 100 high-coverage genomes from a pedigreed captive baboon colony. *Genome research*, 29(5), 848–856. <https://doi.org/10.1101/gr.247122.118>

Rogers, J., Raveendran, M., Harris, R. A., Mailund, T., Leppälä, K., Athanasiadis, G., Schierup, M. H., Cheng, J., Munch, K., Walker, J. A., Konkel, M. K., Jordan, V., Steely, C. J., Beckstrom, T. O., Bergey, C., Burrell, A., Schrempf, D., Noll, A., Kothe, M., Kopp, G. H., ... Baboon Genome Analysis Consortium (2019). The comparative genomics and complex population history of *Papio* baboons. *Science advances*, 5(1), eaau6947. <https://doi.org/10.1126/sciadv.aau6947>

Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American journal of physical anthropology*, 174(3), 407–417. <https://doi.org/10.1002/ajpa.24186>

Samuels, A., & Altmann, J. (1986). Immigration of a *Papio anubis* male into a group of *Papio cynocephalus* baboons and evidence for an *anubis-cynocephalus* hybrid zone in Amboseli, Kenya. *International Journal of Primatology*, 7, 131–138.

Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., Alessandri, L., Arienzo, I., De Angelis, F., Rolfo, M. F., Skeates, R., Silvestri, L., Beckett, J., Talamo, S., Dolfini, A., Miari, M., ... Scheib, C. L. (2021). Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Current biology : CB*, 31(12), 2576–2591.e12. <https://doi.org/10.1016/j.cub.2021.04.022>

Schildkrout, E. (1989). Art as Evidence: A History of the American Museum of Natural History African Ethnology Collection. In Art/Artifact, ed. A. Danto, S. Vogel, et. al. New York: The Center for African Art, pp. 153–160

Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 759.

Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., Storer, J. M., Kuhlilm, M., Fontser, C., Seshadri, L., Bergey, C. M., Burrell, A. S., Bergman, J., Phillips-Conroy, J. E., Shiferaw, F., Chiou, K. L., Chuma, I. S., Keyyu, J. D., Fischer, J., Gingras, M. C., ... Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science (New York, N.Y.)*, 380(6648), eabn8153. <https://doi.org/10.1126/science.abn8153>

Sugawara, K. Sociological study of a wild group of hybrid baboons between *Papio anubis* and *P. hamadryas* in the Awash Valley, Ethiopia. *Primates* 20, 21–56 (1979).

Sugawara, K. (1988). Ethological study of the social behavior of hybrid baboons between *Papio anubis* and *P. hamadryas* in free-ranging groups. *Primates*, 29, 429–448.

Tung, J., Charpentier, M. J. E., Garfield, D. A., Altmann, J., & Alberts, S. C. (2008). Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Molecular Ecology*, 17, 1998–2011.

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, 2019, pp. 314-324. <https://doi.org/10.1109/IPDPS.2019.00041>

Wall, J. D., Schlebusch, S. A., Alberts, S. C., Cox, L. A., Snyder-Mackler, N., Nevonen, K. A., Carbone, L., & Tung, J. (2016). Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of Admixture in wild baboons. *Molecular ecology*, 25(14), 3469–3483. <https://doi.org/10.1111/mec.13684>

Wandeler, P., Hoeck, P. E., & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in ecology & evolution*, 22(12), 634–642. <https://doi.org/10.1016/j.tree.2007.08.017>

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American journal of physical anthropology*, 150(1), 133–140. <https://doi.org/10.1002/ajpa.22185>

Chapter 3 - *Theropithecus gelada* and the intergenic hybridisations with *Papio*

This chapter has been entirely extracted from Caldon et al. 2024 “Gelada genomes highlight events of gene flow, hybridisation and local adaptation that track past climatic changes” published in Molecular Ecology under licence CC BY 4.0.

The text of the article has been adapted to the appropriate format for the thesis. All the figures have also been modified from those in the original article. Analyses not conducted by the author of this thesis have not been included in the materials and methods or results sections, but are referenced in the discussion, with proper citation of the article. Sections, where the work of the author of this thesis relied on others' contributions, will be duly noted in the methods section. Finally, supplementary figures and tables not created by the thesis author, or supplementary figures that are not necessary but cited in the following text, will not be included in the chapter. Instead, reference will be made to them in their original form within the original article, for simplicity the full paper and the related useful Supplementary Material are included in Appendix 1.

The only addition is a novel set of unpublished results in section 3.3.3 that concerns D-statistics with more olive baboon populations, which were performed after the publishing of the paper and alongside the work for Chapter 2 of this thesis. This addition is useful for exploring gene flow between *Theropithecus* and *Papio anubis* not as a species (as in the paper) but as different and heterogenous populations (as Chapter 2 results suggest).

Preliminary results from this chapter were been presented as a poster at SMBE (Society for Molecular Biology and Evolution) congress in 2023 at Ferrara, Italy.

3.1 - Introduction

Theropithecus gelada is the only extant species belonging to the primate genus, *Theropithecus* (FIGURE 3.1; Delson, 1993). This species occupies a unique ecological niche, the Afroalpine grasslands of the Ethiopian Highlands, and is specialised for altitudes ranging from 2000 to 4500 m (Belay & Shotake, 1998). Geladas are the sole graminivorous primates, consuming a grass-based diet similarly to ungulates (Iwamoto, 1993a). Their diet comprises primarily graminoid leaves (up to 90% of the ingested food; Dunbar, 1977). The second most common dietary

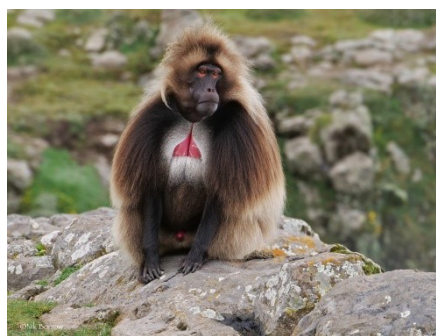


FIGURE 3.1 - *Theropithecus gelada* (male). Photo by Nik Borrow. Image via Flickr (<https://www.flickr.com/photos/nikborrow/48899037528>) licensed under Creative Commons (CC BY-NC 2.0).

component is underground food (UF: tubers, roots and corms), a nutritional fall-back resource rich in starch and mostly consumed when the grass is scarce as the result of decreased rainfall (Iwamoto, 1993a, 1993b; Jarvey et al. 2018).

On the basis of morphological traits, two major subspecies have been described: *T. g. gelada* (Rüppell, 1835) and *T. g. obscurus* Heuglin, 1863. The two subspecies are usually associated respectively with the Northern region and the central region of the Northern Ethiopian plateau (de Beaux, 1925; Gippoliti, 2010; Girmay & Tesfay, 2020). However, the absence of clear-cut differences, particularly when considering the variation in coat pigmentation, and the apparent absence of a major barrier between the two taxa has led to scepticism about the taxonomic validity of the subspecies classification (Yalden et al. 1976, 1996). Mitochondrial DNA (mtDNA) phylogenetic analyses are broadly congruent with the Northern-Central subdivision first proposed by de Beaux (de Beaux, 1925; Shotake et al. 2016; Zinner et al. 2018). The geographical range of modern-day geladas was extended in 1990 when a population was found South of the Great Rift Valley in the South-Eastern plateau, in the Arsi mountains (Mori & Belay, 1990). In 2016 the subspecies *Theropithecus gelada arsi* Shotake Saijuntha Agatsuma et Kawamoto, 2016, was tentatively proposed for this clearly isolated population on the basis of coat coloration and molecular data (Belay & Shotake, 1998; Gippoliti, 2010; Mori & Belay, 1990; Shotake et al. 2016).

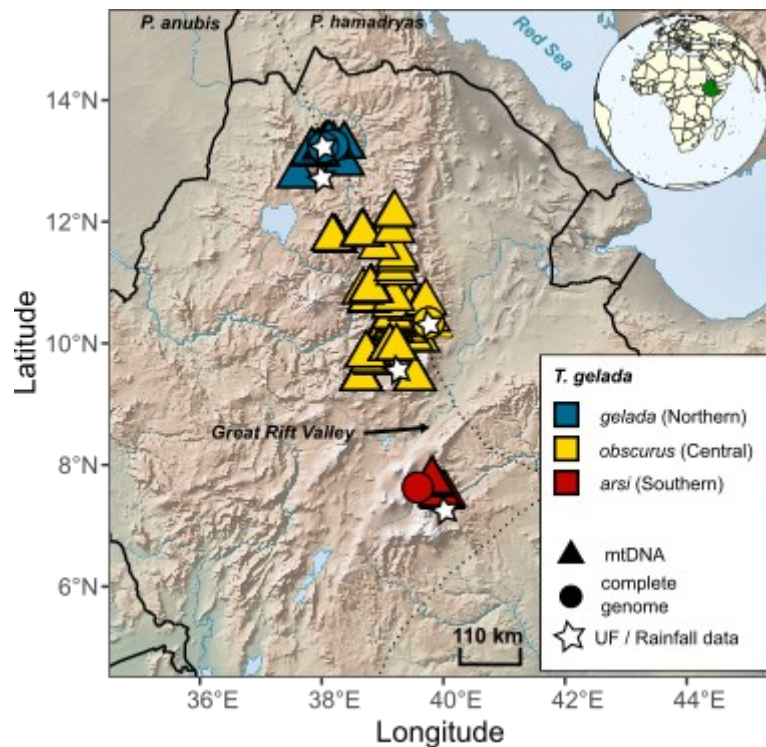


FIGURE 3.2 - Geographical origin of *Theropithecus gelada* samples analysed in this work. Colours refer to source populations (Northern, Central and Southern) and shapes identify the data available for the sample (Complete genome sequence and mitochondrial DNA only). The dotted line provides an approximate indication of the distribution of *P. hamadryas* (right) and *P. anubis* (left) in the area of interest, as indicated in Sørensen et al. (2023). The stars provide the location of provenience of the five datasets with UF consumption and rainfall data (from top to bottom: Northern: Sankaber (Jarvey et al., 2018) and Gich (Woldegeorgis & Bekele, 2015), Central: Guassa (Fashing et al., 2014) and Wollo (Kifle & Bekele, 2021), and Southern: Indetu (Abu et al., 2018)). Prepared using Natural Earth. Free vector and raster map data @natural earth data. com. Figure and caption from Caldon et al. 2024.

Given the debate about the geographical distribution of the three gelada subspecies and concerns about their taxonomic validity, in this work, we will name the three gelada demes as Northern, Central and Southern geladas, as done by others in previous investigations (Chiou et al. 2022; Trede et al. 2020; Zinner et al. 2018) acknowledging that some degree of overlap is present between these geographically described units and the *T. g. gelada*, *T. g. obscurus* and *T. g. arsi* taxa (FIGURE 3.2; Gippoliti, 2010; Shotake et al. 2016).

Ethiopian Rift Valley (Central and Southern geladas), with the two groups being characterised by similar levels of within-group variation (Belay & Shotake, 1998). Notably, the mitochondrial lineages of the Southern geladas of the Arsi mountains are identified either as basal to all the other geladas (Central and Northern; Shotake et al. 2016) or are placed within the Northern variation, sister to a Northern subclade (Zinner et al. 2018). Despite these differences in tree topologies, the investigations conducted by Zinner et al. (2018) and Shotake et al. (2016) are concordant in supporting a relatively recent Time to the Most Recent Common Ancestor (TMRCA) of gelada mtDNA (670–400 thousand years ago, kya), and in showing Southern geladas as characterised by the lowest mtDNA diversity across the three gelada populations. Interestingly, no lineages are shared between geladas living in the three regions (with the exception of a likely mislabelled sample, as suggested by Zinner et al. (2018)), suggesting deep population structure among the three groups, at least for the maternal lineages. Microsatellite data from a subset of the samples analysed by Zinner et al. (2018) confirmed a three-deme structure for geladas (Trede et al. 2020), the genetic distance, measured using F_{ST} , between Southern geladas and the other two populations being two-three times larger than the distance between Central and Northern (Trede et al. 2020). More recently, whole genome sequence data of individuals from two regions in the Northern (Sankaber and Chenek), one in the Central (Guassa) and five US zoos were released (Chiou et al. 2022). The published preliminary population genomic analysis of this dataset confirmed the subdivision between Northern and Central geladas, zoo specimens (ultimate origin of these animals or their ancestors was unknown) showing mostly a Central gelada ancestry component. The Northern-Central separation also emerged at karyotypic level as a different number of chromosomes were suggested for individuals sampled in the two regions (42 and 44 respectively). Centric fission of chromosome 7 was detected in Northern gelada, resulting in two acrocentric chromosomes, 7a and 7b, and $2n = 44$ karyotype. Such a macroscopic genomic difference was interpreted as evidence for putative incipient speciation as hybrids between lineages with heterozygous karyotypes might have reduced reproductive success (King, 1995). Genome-wide data were also used to estimate how the effective population size of Northern and Central geladas changed in the past. Both showed an almost steady decrease over the last million years; notably, Northern gelada and Central gelada diverged in their pattern 500 kya, suggesting this as the initial temporal point of separation of the two groups (Chiou et al. 2022).

The Ethiopian plateau is part of the East African Rift System, an area profoundly affected by environmental instability during the late Quaternary (Foerster et al. 2012). Geochemical, physical and biological indicators testify that Ethiopia's climate regime was controlled by the intensification of the dry–wet–dry cycles during the last glacial period, potentially linked to the change in orbitally forced radiation (Trauth et al. 2003). Species response to such changes usually includes habitat tracking towards different latitudes, altitudes or the population concentrated within refugia (Keppel et al. 2012; Raia et al. 2012). The gelada is probably no exception since the species is currently confined to the Ethiopian Highland at high altitudes whereby the

three recognised populations occupy different areas of the plateau (Keppel et al. 2012; Raia et al. 2012).

In addition, hybridisation, another process potentially influenced by climatic variation, has been suggested between geladas and members of the genus *Papio*, possibly having occurred more than once, on the basis of mobile elements being shared across the genomes of baboon species and geladas (Walker et al. 2019). However, besides the rejection of a phylogenetic tree that excluded gene flow between geladas and *Papio* (Santander et al. 2022), no formal testing for hybridisation event(s) has been performed so far. Mating events between geladas and baboons and the possible presence of hybrids have been reported both in captivity and in the wild (Dunbar & Dunbar, 1974; Jolly et al. 1997; Markarjan et al. 1974).

Overall, the emerging consensus based on DNA supports at least three evolutionary units for *T. gelada*, with a robust parallelism between genetics and geography, possibly the result of their adaptation to the montane grassland environment (Chiou et al. 2022). However, the impact of climatic changes on geladas current geographical distribution and past population dynamics has been poorly explored, particularly considering the intense climatic oscillations of the Late Pleistocene. Here we integrate population genomic data with explicit modelling of species habitat suitability to better clarify the linkage between species ecological preferences and demographic dynamics, with the aim of generating an integrated reconstruction of the evolutionary history of *Theropithecus gelada*.

As is often the case in biological conservation, reaching a consensus on taxonomy and conservation units by integrating as many different data sets as possible (Gippoliti, 2020; Groves et al. 2017) may be crucial for immediate management actions for *Theropithecus*. This may include the identification of priority sites for research and conservation in Ethiopia, the recognition of the significance of ex-situ population programs and the future planning for re-stocking and/or reintroduction actions. We note that while the conservation status of the overall species is Least Concern (LC; Gippoliti et al. 2019) the geographical features of the Ethiopian plateau can endanger the survival of populations due to isolation and small census size, as well exemplified by Southern geladas.

3.2 - Materials and Methods

3.2.1 - Sampling and Samples Dataset

Gelada blood samples from different locations in Ethiopia were collected by two of the authors (Gurja Belay and Takayoshi Shotake) during 1978–1997 (more details can be found in Belay & Shotake, 1998; Shotake et al. 2016). The guidelines for field research of the Kyoto University Primate Research Institute and the rules and regulations of the Ethiopian government at the time were followed during blood collection (Belay & Shotake, 1998). From this collection, DNA samples from the Arsi (Southern) population were evaluated for their compatibility with whole-genome sequencing, and two samples, one from Gado-Goro (A28) and one from Lega Aba Boge-Lega Adem (A10), were selected to be whole-genome sequenced. Sample A28 was sequenced to a depth of 30×, while sample A10 reached a median coverage of 12×, the sequencing being completed by EuroFin Japan (Supplementary Table 3). In addition, we sequenced a gelada

sample provided via the European Association of Zoos and Aquaria (EAZA) by the Wilhelma Zoo Stuttgart and processed by the EAZA biobank hub at the Leibniz Institute for Zoo and Wildlife Research (HJZ14-00788), here referred to as TGLZ1 (the ultimate geographical origin of all captive animals included in this study and their ancestors is unknown). The sample TGLZ1 was sequenced at the Centro Nacional de Análisis Genómico in Barcelona, reaching a final coverage of 33.65× (Supplementary Table 3). We also generated whole genome sequence data from a sample of *Lophocebus aterrimus* provided by the Gaia Zoo Kerkrade and processed by the EAZA hub at Copenhagen Zoo. This sample was sequenced at the Centro Nacional de Análisis Genómico in Barcelona with a mean coverage of 34.98× (Supplementary Table 3). The taxonomic position of the genus *Lophocebus*, sister to *Papio*, makes it possible to highlight *Papio*-specific evolutionary dynamics.

In this study, we also included all the samples of *T. gelada* and one of *Papio hamadryas* (hamadryas baboon) from Chiou et al. (2022), one sample of *Papio anubis* (olive baboon) from Sørensen et al. (2023) and one sample of *Mandrillus leucophaeus* (drill) recovered from PRJNA785018 BioProject in NCBI (Supplementary Table 3). We focused our attention on the two *Papio* species present in Ethiopia (*P. anubis* and *P. hamadryas*) and known to have been in contact with geladas (Dunbar & Dunbar, 1974; Jolly et al. 1997). In this study, we used the term hybridisation to refer to interbreeding between *T. gelada* and *Papio*, at any point in time in their evolutionary history, and admixture for gene flow between gelada populations.

3.2.2 - Genome Mapping

Reads were mapped with BWA-MEM2 v.2.2.1 (Vasimuddin et al. 2019), to the reference assembly genome Tgel_1.0 (Chiou et al. 2022). Since there was no Y chromosome in this reference genome, we added it from the *Papio* reference assembly genome Panubis1 (Batra et al. 2020). Reads from (Chiou et al. 2022) were trimmed with fastp v.0.23.2 (Chen et al. 2018) in order to remove Nextera adapters. We then marked the duplicate reads with Picard MarkDuplicates version 2.8.1 (<http://broadinstitute.github.io/picard/>) and filtered the result using the options -q 10, -F 1292 and -f 2 of samtools view (Li et al. 2009).

To avoid some analyses being influenced by the reference genome (in particular, those involving comparisons with other genera, see later; Günther & Nettelblad, 2019), we additionally mapped *Papio*, *Lophocebus*, *Mandrillus* and a subset of *Theropithecus* samples (11 high-coverage gelada individuals belonging to the four localities of sampling—for the details see Supplementary Table 3) to the *Macaca mulatta* reference genome (Mmul_10; Warren et al. 2020), following the same procedure described above.

3.2.3 - Variant Calling and the Final Dataset

We called the autosomal chromosome variants, using the GATK version 4.2.4.1 (McKenna et al. 2010). For each sample, we generated the gVCF with HaplotypeCaller, and then we made the joint calls via GenotypeGVCFs. The VCF thus obtained was filtered with VariantFiltration using hard filters for SNPs ('QD < 2.0 || MQ < 40.0 || FS > 60.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0'). The final VCF was therefore composed of 70 individuals of *T. gelada* and 9,244,659 variants. Using PLINK v.1.9 (Chang et al. 2015), all sites from the sex chromosomes and with a minor allele frequency (MAF) below 0.05 were removed, keeping only biallelic SNPs filtered using the --snps-only option.

3.2.4 - Autosome Analyses

3.2.4.1 - Heterozygosity, ROH and Kinship

To calculate the proportion of heterozygous sites (HET) we used PLINK --het. This command estimated the observed O(HOM) homozygous genotype counts for each sample and the total number of variable sites (N_Sites), from which we derived HET per individual with the formula:

$$HET = \frac{N_{Sites} - O(HOM)}{N_{Sites}}$$

We estimated the number, the total length and the average length of runs of homozygosity (RoH) with PLINK --homozyg and the option --homozyg-snp 25 --homozyg-kb 200 --homozyg-window-snp 10. The degree of kinship among samples was tested up to the third degree using KING v.2.3.1 with --kinship -- degree 3 flags on the PLINK bed file (Manichaikul et al. 2010).

3.2.4.2 - PCA

We performed a preliminary survey of *T. gelada* population structure by running a principal component analysis (PCA) using the smartpca function implemented in EIGENSOFT software 8.0.0 (Patterson et al. 2006) using default settings.

3.2.4.3 - D-statistics

We performed the D-statistics (ABBA-BABA test) using qpDstat (Patterson et al. 2012), with default parameters, to test for the occurrence of any imbalance in allele sharing between different populations and species in the form D(H4,H3;H2,H1) including drill (*M. leucophaeus*) as an outgroup (H4). We firstly tested whether the two Northern populations of Sankaber and Chenek (as H1 and H2) were symmetrical in their sharing with other gelada groups (Central or Southern as H3) and what was the degree of allelic sharing between the three gelada taxa (Northern, Central, Southern alternatively placed as H1, H2 or H3). Then we tested the allelic sharing of these gelada populations with *Papio* species and *Lophocebus* (Dunbar & Dunbar, 1974; Jolly et al. 1997) in the form D(Drill, Papio; Gelada2, Gelada1), D(Drill, Lophocebus; Gelada2, Gelada1), D(Drill, Gelada; Papio2, Papio1) and D(Drill, Gelada; Lophocebus, Papio), where ‘*Papio*’ were alternatively *P. anubis* or *P. hamadryas* genomes, ‘*Lophocebus*’ was the genome of *L. aterrimus* and ‘Gelada’ was a genome from one of the three gelada populations.

Finally, we tested the olive baboon clusters obtained in Chapter 2 as H2 and H1 and the Southern gelada (as H3).

The test considers the ancestral alleles (A) and the derived alleles (B) in the genomes of the four taxa considered. Without gene flow, the two conflicting allelic patterns (ABBA and BABA) are equally frequent, and the values of D are not significantly different from zero. In the presence of gene flow an excess of one of the two allelic patterns occurs, which produces a deviation of the D from zero: positive D (excess of BABA) indicates the introgression between H1 and H3, and negative D (excess of ABBA) points to the gene flow between H2 and H3. The estimates were considered significant if |Z|-score >3.

The D-statistics analysis involving the genera *Papio* and *Lophocebus* was conducted using reads mapped on the Mmul_10 reference genome, to avoid a reference bias.

3.2.4.4 - Admixture graphs

We explored the relationships and the gene-flow events between the three gelada taxa and the species of the genus *Papio* (*P. anubis* and *P. hamadryas*) and *L. aterrimus* using `find_graphs` in R package `admixtools2` (Maier et al. 2023). The R command `find_graphs` was run with parameters `stop_gen = 10,000`, `stop_gen2 = 100`, `plusminus_generations = 20`, `eject_f4z = 2`, `diag = 1e-04`, `numstart = 1000`, `lsqmode = FALSE`, `resid = F`, `num_admix` between 1 and 5 and *M. leucophaeus* as an outgroup; the same command was run 10 times per number of admixture events in order to obtain the graph with the score as close as possible to the global optimum. We kept, for each edge, the three trees with the best score; we reported in the main text the tree with the greatest number of admixture events and no contributions equal to 0%. The admixture graphs analysis was conducted using reads mapped on the macaque reference genome Mmul_10, to avoid reference bias.

3.2.4.5 - PSMC

Demographic changes (effective population size history) over time were explored using the pairwise sequentially Markovian coalescent (PSMC) model (Li & Durbin, 2011). As recommended, we generated diploid sequences using `bcftools` (Li, 2011); sites with coverage greater than double the average genome-wide coverage or less than one-third were excluded. We also excluded the repetitive regions in the genomes, indicated on the NCBI website (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_003255815.1). We run PSMC with default settings on autosomal chromosomes, using a mutation rate of 0.5×10^{-8} per base pair per generation and a generation time of 11.67 years as in Chiou et al. (2022). Finally, we performed bootstrapping for each individual with 10 replicates following the instructions described on the GitHub webpage manual (<https://github.com/lh3/psmc>).

In a similar manner, this software can also be used to explore the last time of contact between populations by creating artificial F1 hybrids (hPSMC; Cahill et al. 2016). Following the procedure described on the GitHub webpage manual (<https://github.com/jacahill/hPSMC>), we generated from the bam a haploid sequence in fasta format of the X chromosome for each male individual. We then combined the fasta sequences from two different individuals into a single `.psmcfa` file, and then we applied PSMC with default settings. We use a mutation rate of 0.419×10^{-8} mutations per base pair per generation. It became necessary to recalculate the autosomal mutation rate, as the X chromosome, due to its presence in different numbers in male and female individuals, results in a different mutation rate. Therefore, the parameter was obtained by converting the autosomal mutation rate into X mutation rate as described by Li and Durbin (2011), using the male-to-female mutation ratio of *Rhesus macaque* (2.9; Wilson Sayres et al. 2011). Finally, we performed bootstrapping for each hybrid as described above.

To test if the estimates obtained with hPSMC are influenced by the selected reference, we performed the procedure described above using reads mapped to the two reference genomes (Tgel_1.0 and Mmul_10).

3.2.5 - AMY genes (Amylase)

These analyses were conducted in collaboration with Giacomo Mutti, Jordi Morata and Jean-Rémi Trotta.

We further explored differences across gelada populations by examining gene content variation. Previous scans of gelada genomes identified gene families significantly expanded compared to *Papio*, highlighting biological processes such as cell proliferation, hypoxia response, and immune function (Chiou et al. 2022). Copy number variation (CNV) has been shown to contribute significantly to phenotypic diversity within species (Almarri et al. 2020; Montinaro & Capelli, 2020; Sudmant et al. 2015).

The gelada genomic reads mapped to Tgel_1.0 were processed via Control-FREEC to identify variations in copy number of genomic regions across the three populations of geladas (Boeva et al. 2012). The number of copies estimated for each locus was reported for all the individuals, clustered according to the population of provenance (Northern, Central and Southern). For each expanded region the number of individuals bearing more than two copies was counted to estimate their overall occurrence and the frequency within each population. Annotated genes overlapping these regions were also reported. KEGG enrichment analyses were conducted via the clusterProfiler v.4.0.5 R package (Wu et al. 2021).

We used Control-FREEC to identify CNV in the gelada genome (Tgel_1.0) across the Northern, Central, and Southern populations (Boeva et al. 2012). For each expanded genomic region, we counted the number of individuals with more than two copies and reported overlapping annotated genes. KEGG enrichment was performed using the clusterProfiler R package (Wu et al. 2021).

To verify the presence of amylase genes in geladas, we searched the reference genome for matches to human salivary and pancreatic amylases (AMY1A, 2A, and 2B). Two genes were identified, both annotated as pancreatic (XP_025215097.1 and XP_025215076.1) on chromosome 1. The two genomic regions hosting these genes were identified on chromosome 1: one between positions 131757727-131766740 (LOC112607926) and the other spanning positions 131799832-131808869 (LOC112607935). Amylase gene copy number was further verified in geladas using the average coverage across the two gelada amylase genomic regions, normalising it by the average coverage across chromosome 1, in each of the gelada genomes.

We reasoned that in the presence of positive selection operating on a given haplotype, the areas surrounding the locus under selection should be characterised by a decrease in genetic diversity, the size of the chromosomal region affected by this pattern being shaped by the time since the selective pressure has been in place. In order to test the hypothesis of positive selection for the duplication of the amylase locus LOC112607926 in Southern gelada, we estimated the degree of heterozygosity present in the regions surrounding the duplication in all gelada genomes.

We partitioned the 10 Mb region centered on the AMY genes in windows of 10kb; starting from the VCF, for each window, we calculated the number of heterozygous sites per individual using VCFtools --het. We then calculated the average for each window for each population and normalised the results by the mean of the variable sites across the entire chromosome. Starting

from the windows containing the loci, we estimated the cumulative sum of the heterozygosity in each window. In order to take into account the differences across populations in the overall degree of heterozygosity, the cumulative sums were normalised by the average of heterozygosity estimated in all the 10kb-windows of chromosome 1. Finally, we calculated the average for each population, including zoo individuals in the Central population.

3.2.6 - Variation in Underground Food Consumption and Rainfall

Amylase expression has been linked to variation in copy number of amylase genes as the result of differences in dietary starch prevalence across human populations, dogs and pigs (Axelsson et al. 2013; Groenen et al. 2012; Pajic et al. 2019; Perry et al. 2007). Seasonal rainfall variation has been reported to affect grass availability and Underground Food (UF; food rich in starch whose digestion is facilitated by amylases) consumption by geladas in Sankaber, grass availability significantly related to the cumulative rainfall in the previous 1–3 months (Jarvey et al. 2018). To further investigate the UF dietary content across gelada populations, we extracted from previous publications dietary data of geladas in different locations (FIGURE 3.2). Only sites with at least 1 year of continuous data collection were included, which resulted in four datasets collected over different periods: 18 months in Kosheme, Wollo (Central; data collected from May 2015 to May 2016, plus July, September, Nov 2016 and January 2017; Kifle & Bekele, 2021), 15 months in Guassa (Central; February 2007 to April 2008; Fashing et al. 2014), 12 months in Sankaber (Northern; February 2015 to January 2016; Jarvey et al. 2018) and 12 months in Gich (Northern; May 2013 to April 2014; Woldegeorgis & Bekele, 2015). We additionally included data from the Southern gelada population of Indetu, the only site with dietary data available South of the Great Rift Valley (Abu et al. 2018). Indetu comprised a set of continuous observations over 7 months extending from August 2010 to February 2011, followed by a gap in March 2011, and two additional data points for April and May 2011. In order to provide a comparison between the five datasets we initially estimated the monthly dietary content of UF (here measured as feeding time spent collecting and processing, including chewing and swallowing- dedicated to any underground part of grass and herbs, including any unclassified underground food (Jarvey et al. 2018)) over the 6-month period shared by all five datasets (August–January). We then repeated the same estimate using a sliding window of 7 months, the longest continuous set of observations in Indetu, in the other four datasets using only data collected in subsequent months with no gaps (number of windows per population: seven for Wollo, eight for Guassa, five for Sankaber and Gich). Given the link between rainfall, grass availability and UF consumption (Jarvey et al. 2018), we collected data on monthly total rainfall for the period spanning the years sampled for dietary consumption (2006–2017) for the five locations (as measured within a 20 squared km) from CRU-TS 4.06 (Harris et al. 2020) downscaled with WorldClim 2.1 (Fick & Hijmans, 2017; Supplementary Table 4). Spearman correlation coefficient r_s and associated p-value between UF consumption and rainfall for the full dataset were calculated (total number of included months across locations: 66). We repeated this analysis by alternatively considering rainfall over 1, 2 or 3 months earlier than the month taken into consideration for UF consumption (Jarvey et al. 2018), as well as including in the analysis the rainfall in the month of interest for UF consumption. Evidence of significant variation in the amount of UF consumption and rainfall across the five locations was tested using one-way ANOVA.

3.3 – Results

3.3.1 - *T. gelada* Population Structure and Genomic Variation

For the first time, the whole genome of individuals sampled from the gelada population located South of the Rift Valley (*T. g. arsi*) was investigated. The PCA results confirm the presence of structure within the species. The newly analysed Southern population of *T. gelada* is distinct from the two previously reported populations (Northern and Central), as suggested by autosomal STRs and mitochondrial DNA (Shotake et al. 2016; Trede et al. 2020; Zinner et al. 2018; FIGURE 3.3A; Figure 2a–b in Appendix 1). The Northern population is separated from the other two along principal component 1 (PC1) which summarises 35.62% of the total diversity, while the Southern population separates from the others along PC2 which comprises 4.85% of the variation. As shown by Chiou et al. (2022), individuals from zoos belong to the Central population as our sample from the zoo in Stuttgart (TGLZ1).

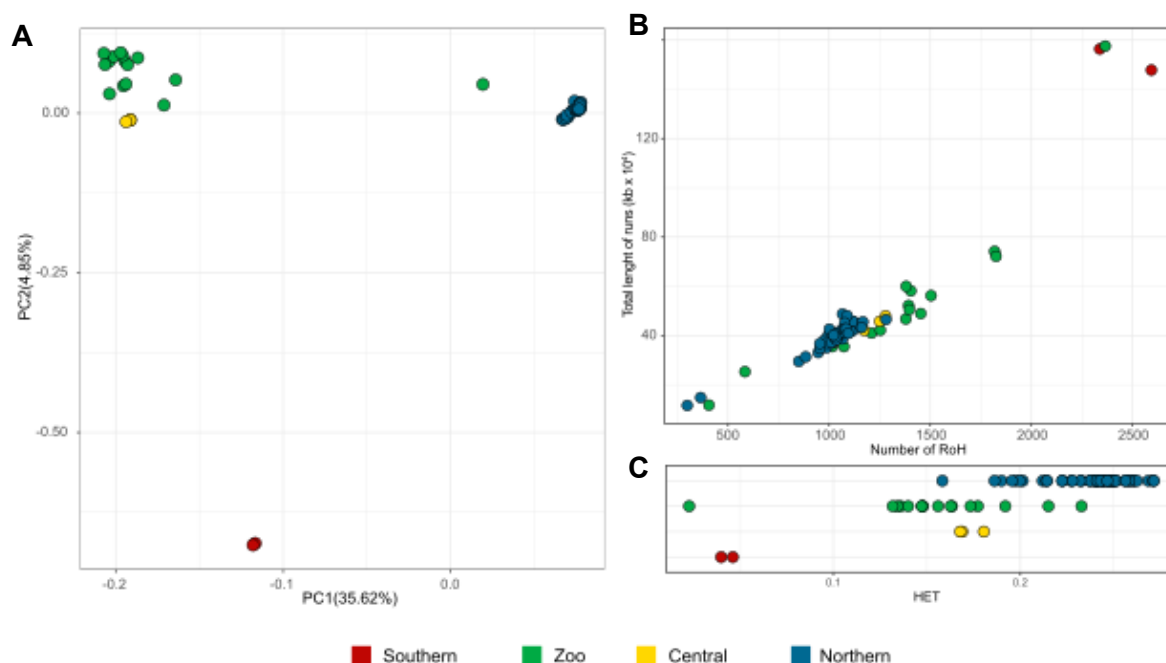


FIGURE 3.3 - Population structure of *Theropithecus gelada*. A) Principal component analysis of autosomal SNPs. B) Intra-individual variation, reported for each sample as number of runs of homozygosity (RoH, x-axis) and total length of RoH (y-axis). C) Autosomal SNPs proportion of observed heterozygous sites (HET) in investigated samples partitioned by population. Figure (partially modified) and caption from Caldon et al. 2024.

Compared to other geladas, Southern individuals are clear outliers for the proportion of observed heterozygous sites, total length and number of RoH. The only exception was SAMN20949854, a zoo sample that showed a lower heterozygosity, a greater number and a longer total length of RoH than other zoo samples (FIGURE 3.3BC) whose coverage at 1.29× is also much lower than the rest of the samples here analysed (mean 14.5×; Supplementary Table 3). We also tested the relationship between individuals up to the third degree and we noted that three pairs of geladas individuals from the Northern (Sankaber) population have a first-degree relationship; nine pairs have a second-degree relationship, including the two from the

Southern population (sampled in different locations); and another nine pairs have a third-degree relationship, including two involving pairs of captive animals, one being TGLZ1 (Figure S2 in Appendix 1). It is worth noticing here that for these estimates the allele frequencies consider all the individuals here analysed. As such this might result in artificial closer affinities if the source populations have substantial lower diversity, as possibly the case for the Southern population.

3.3.2 - Demographic History of Gelada Populations

The reconstruction of demographic changes across time was performed using the PSMC software (Li & Durbin, 2011; FIGURE 3.4). Results indicate that all three gelada populations experienced decreasing population size and/or connectivity over the last half a million years, differently from what was experienced by two geographically close Ethiopian Papionins, *Papio anubis* and *P. hamadryas* (Figure S3 in Appendix 1). Across geladas, the magnitude of this reduction was more pronounced for the Southern population (Chiou et al. 2022). The size of the three populations remained small over the last 50,000 years.

In order to have an indication of the timing of the most recent events of gene flow between different gelada populations, we reconstructed the demographic curves of artificial F1 hybrids of chromosome X through hPSMC (Cahill et al. 2016; FIGURE 3.4). In accordance with coalescent theory the absence of coalescent events is indicative of extremely large populations, which results in the curves exponentially increasing at the time of the latest MRCA. The curves for the hybrids between gelada populations show this pattern at 10–20 kya, suggesting this was the time of last genetic exchange between groups. We observed the same results when the hPSMC analyses were replicated making use of the reads mapped to the Mmul_10 reference genome (Figure S4b in Appendix 1).

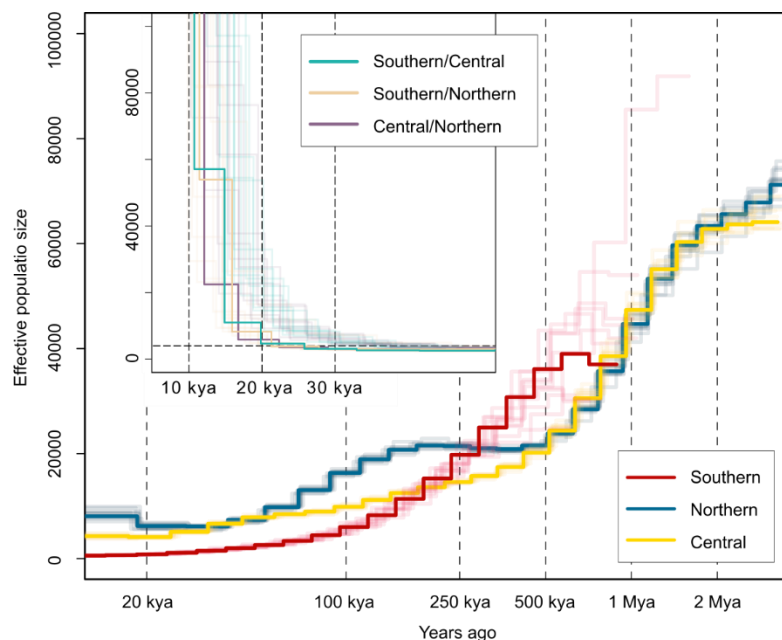


FIGURE 3.4 - Demographic history of gelada populations. Changes in effective population size across time estimated using PSMC. SAMN20949878, SAMN20949865 and A28 genomes were analysed for Northern Central and Southern geladas respectively. X-axis on logarithmic scale. Colours as in Figure 1. Inset: hPSMC coalescence curves of artificial X chromosome hybrids of the same three samples, mapped on Tgel_1.0 reference genome. Figure and caption from Caldon et al. 2024.

We then used D-statistics to further investigate patterns of gene flow across gelada populations by alternatively placing genomes of geladas from the three populations in position H1, H2 or H3. A gene-flow signal between Southern and Central geladas was highlighted when compared to the two Northern populations of Sankaber and Chenek (Figure S4a in Appendix 1), and the two Northern populations showed a signal of genetic exchange with the Central than with the Southern population ($D = 0.1199$, $Z = 16.952$ and $D = 0.1206$, $Z = 16.627$; Figure S4a in Appendix 1). We also noted that the two Northern groups (Sankaber and Chenek) exhibited similar genetic exchange patterns with the Central and the Southern populations ($D = 0.0021$, $Z = 0.613$ and $D = 0.0012$, $Z = 0.301$ respectively); as such, we merged individuals from two Northern groups into a single population for subsequent analyses (Figure S4a in Appendix 1).

3.3.3 - *Theropithecus-Papio* Hybridisation

Following previous suggestions of possible *Papio* and *Theropithecus* hybrids based on molecular data and observations in the field (Dunbar & Dunbar, 1974; Walker et al. 2019), we tested for patterns of shared alleles across the two genera using the D-statistics. We initially tested for any gene-flow signal between gelada and *Papio* genomes when compared to *Lophocebus*. Being sister clades, in the absence of gene flow, *Papio* and *Lophocebus* should be similarly related to *Theropithecus* (Perelman et al. 2011). Our results highlighted that all *Papio* genomes exhibited more shared alleles with geladas than *Lophocebus*, suggesting contacts between the two genera (FIGURE 3.5A). In addition, *Papio* species showed no difference in their degree of shared alleles with different gelada populations and vice versa, except for *P. anubis* and Southern geladas among which was highlighted a signal of gene-flow ($D(\text{Drill, Southern, } P. \text{ hamadryas, } P. \text{ anubis})$ Z-score = 3.038; $D(\text{Drill, } P. \text{ anubis, Southern, Central})$ Z-score = -7.916; $D(\text{Drill, } P. \text{ anubis; Southern, Northern})$ Z-score = -7.410; FIGURE 3.5AB). Overall, these results suggest that gelada populations differ in the amount of between species introgression that occurred in their evolutionary history.

Following the results presented earlier in this chapter and in Caldon et al. (2024, Appendix 1), together with those in Chapter 2 on the unsupervised clustering of olive baboon individuals, which increased the number of available olive baboon samples mapped to the Mmul_10 reference genome, we chose to utilise these new samples and clusters to test allelic sharing signals between Southern geladas and the five identified olive baboon clusters. Results from the new D-statistics analysis (FIGURE 3.5), which tested allelic sharing between Southern geladas (H3) and the five olive baboon clusters (H2 and H1), indicate that the Ethiopian (Et) and Central (Ce) clusters exhibit the strongest allele-sharing with Southern geladas. By contrast, no significant allelic-sharing signals were found in the remaining olive baboon clusters (Te, So, Ar).

We investigated the evolutionary relationships among *T. gelada* populations and *Papio* species (and *Lophocebus*) by means of the approach implemented in find_graphs (Figure S5 in Appendix 1). The topology with the best likelihood score, the greatest number of admixture events and no branches with zero drift estimates is presented in FIGURE 3.5C. The graph underlined the occurrence of an ancient contribution of 11% from the *Theropithecus* branch to the lineage leading to *P. hamadryas* and subsequent smaller contributions to *P. anubis* (6%) and *L. aterrimus* (4%) from a population which also contributed, together with a pre-Central group, to Southern geladas (FIGURE 3.5C).

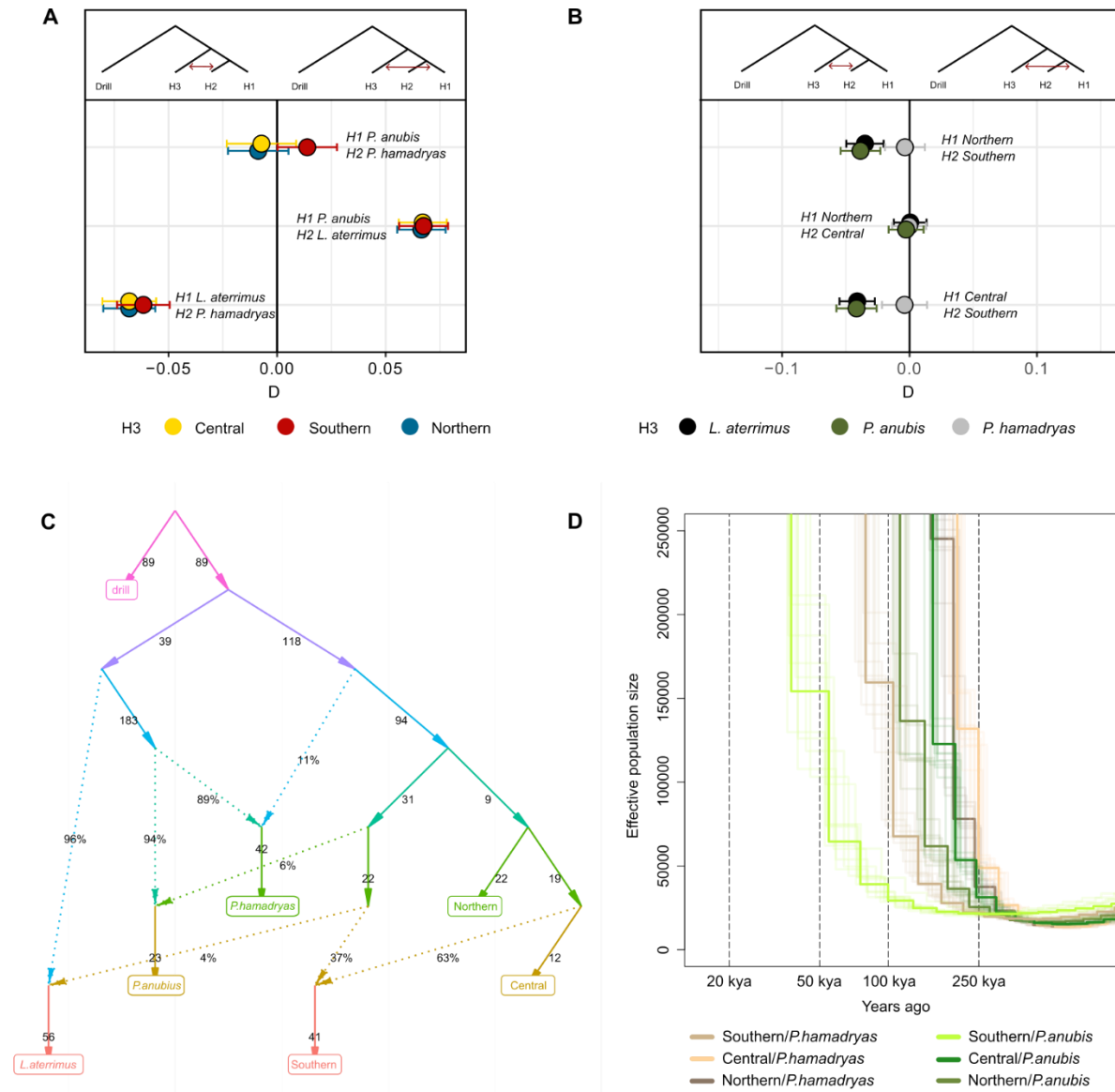


FIGURE 1.5 - Gelada-baboons hybridisation events. (a) ABBA-BABA tests comparing *Theropithecus gelada* populations (H3) and baboons sp. (as H1 and H2); (b) ABBABABA test comparing baboons sp./*Lophocebus aterrimus* (as H3) and *T. gelada* populations (as H1 and H2); bars show the extent of three standard deviations; (c) admixture graphs (built using the R command `find_graphs`), reconstructing relationships and admixture events between Papionines species; dotted arrows indicate admixture, while associated numbers report the amount of each contribution; continuous arrows indicate linear descendants, while associated numbers are indicative of the amount of drift experienced along lineages, given in *f*-statistics units multiplied by 1000; (d) coalescence of artificial hybrids *Papio-Theropithecus* X chromosomes (hPSMC analysis; the gelada samples used are the same in FIGURE 3.4; samples for this analyses were mapped on *Mmul_10* reference genome); x-axis on logarithmic scale. Figure and caption from Caldon et al. 2024.

The generation of *Theropithecus-Papio* hybrid X-chromosomes and their analysis via hPSMC provided the chronological context for the *Papio-gelada* contacts suggested by the D-statistics and highlighted by `find_graphs`. All *P. hamadryas/T. gelada* hybrids generated curves older than 100 kya. Similar curves were generated by *P. anubis/T. gelada* hybrids except for the hybrids involving *P. anubis* and Southern geladas which were more recent (younger than 50 kya; FIGURE 3.5D). We observed the same results when the hPSMC analyses were replicated making use of the reads mapped to the *Tgel_1.0* reference genome (Figure S6 in Appendix 1).

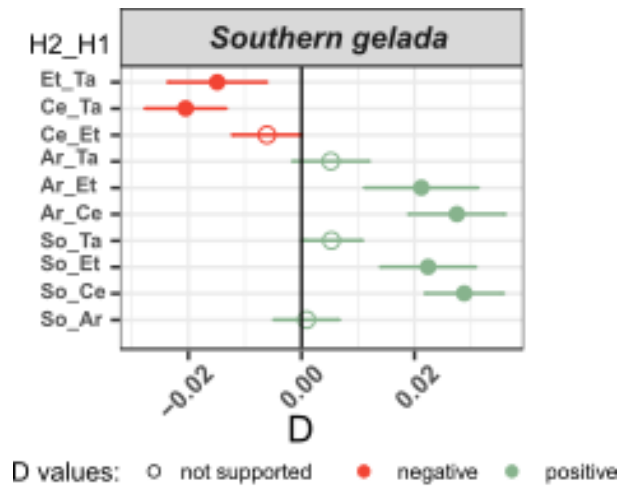


FIGURE 3.6 – Allele sharing between geladas and olive baboons. D-statistics analyses comparing Southern gelada (as H3) and olive baboon clusters (as H1 and H2). The bars show the extension of three standard deviations, the colours refer to negative (red) or positive (green) values of D, and empty dots refer to D values associated with non-supported Z values. For positive values of D, the signal of gene flow is between H3 and H1, while for negative values the signal is between H3 and H2. Abbreviations refer to clusters Ethiopia (Et), Centre (Ce), South (So), Arusha (Ar) and Tarangire (Ta).

3.3.4 - Copy Number Variants

We investigated the pattern of copy-number variants (CNVs) across the three gelada populations by leveraging the population-based genomic data available and newly generated here. A total of 656 variable regions with more than two genomic copies were identified across the 69 available gelada individuals (Table S4 in Appendix 1). The size of the longest expanded regions was up to 5.5 Mb, the vast majority being below 1 Mb (Figure S7a in Appendix 1). Notably, 45% are singletons, appearing only once in the whole gelada dataset (Figure S7b in Appendix 1). 243 of the non-singletons were found only in one of the three populations (FIGURE 3.6A). We considered that local positive selection can drive the increase in the frequency of specific variants in a given population and therefore focused our attention on the set of population-specific CNVs that were present in no less than 25% of individuals in a population (5 in Central and 13 in Northern populations). As only two individuals were analysed for Southern geladas, we considered all the regions uniquely expanded in this population and present in both individuals. In this way, we identified 20, 17 and 10 regions in the Northern, Central and Southern populations, which contained 19, 9 and 6 genes respectively. No significant enrichment in KEGG descriptors was found for the 34 genes across the three populations. We manually inspected the list of genes and noted, among the others, the occurrence of one of the two pancreatic amylase genes (LOC112607926). The two Southern geladas genomes hosted a different number of amylase duplications: three copies were inferred for A28, while four were estimated in A10, the rest of the geladas presenting only two (FIGURE 3.6B). The extent of the duplication was approximately 50 kb in both individuals (spanning positions 131,716,000–131,771,999 and 131,720,000–131,771,999 in samples A28 and A10 respectively; Figure S7c in Appendix 1). Given the reliance of geladas on underground food in the dry season, we decided to further explore the pattern of variation associated with the amylase sequences duplicated in Southern geladas (LOC112607926).

Considering a 10-Mb region surrounding the duplicated amylase locus, Southern geladas showed a decrease in diversity that extended over a chromosome portion much larger than other

populations, up to Mbs in size (FIGURE 3.6C and Figure S7d in Appendix 1). The physical extent of the reduced diversity as the result of positive selection is shaped by recombination events, which are scaled by the number of generations since the selective pressure started. The observed pattern in Southern geladas is compatible with positive selection on the duplicated amylase gene being relatively recent (Sabeti et al. 2002, 2007).

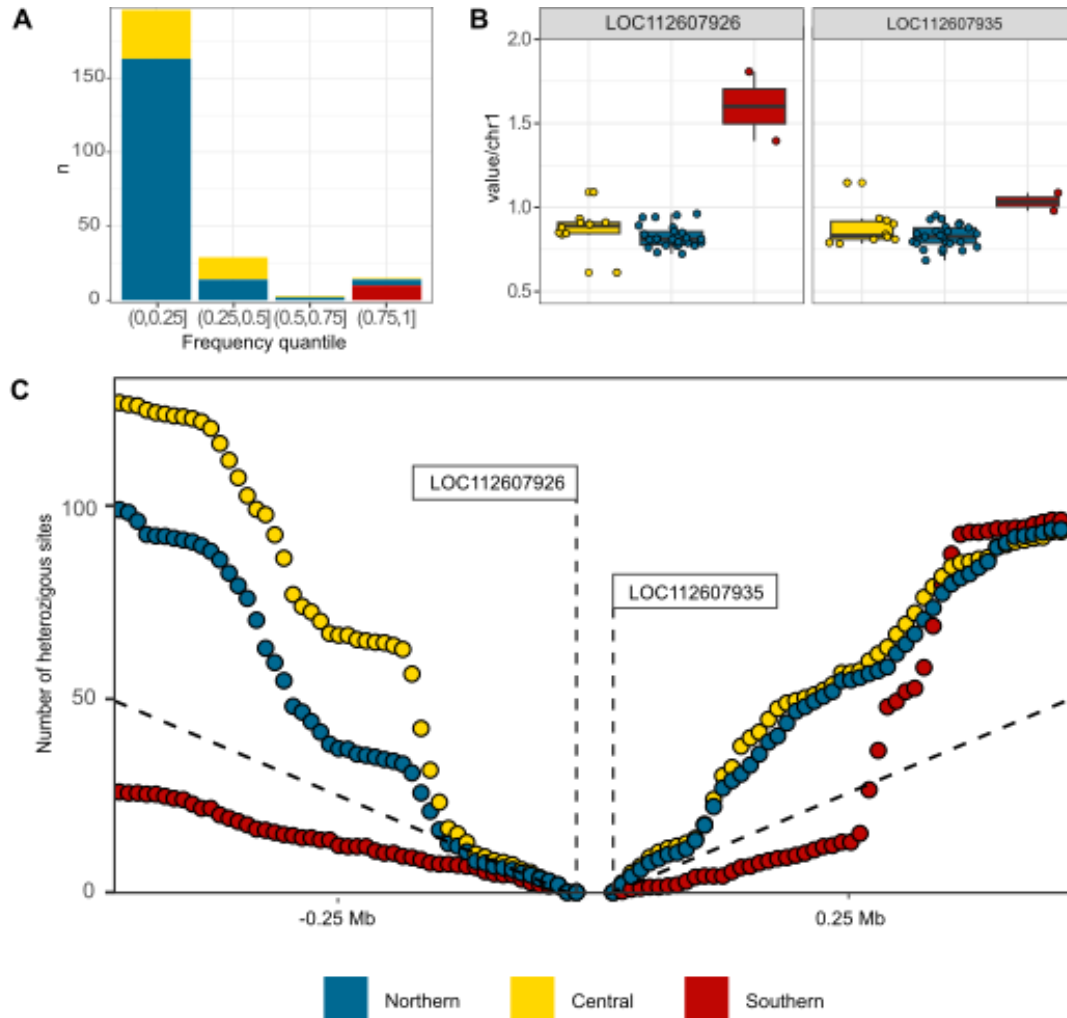


FIGURE 3.6 - CNVs and amylase gene copy number. (a) Population-specific CNVs divided according to their frequency in each population. (b) Normalised genomic coverage over the two loci annotated as pancreatic amylase in geladas. (c) Cumulative number of heterozygous positions in 10-kb windows, starting from the beginning/end of the two geladas pancreatic amylase gene average and extended for 0.5 Mb. Reported is the average across individuals for the indicated population, each individual estimate normalised for the average number of heterozygous positions estimated on chromosome 1 in each individual (see Section 2); dashed lines refer to the expected cumulative curve for 10-kb windows each displaying the average number of heterozygous positions along chromosome 1. Figure and caption from Caldon et al. 2024.

3.3.5 - Variation in Underground Food Consumption and Rainfall

We tested for association between rainfall and UF consumption using the full UF dataset of the five locations investigated here (Sankaber and Gich, Northern; Guassa and Wollo, Central; Indetu, Southern; Fashing et al. 2014; Jarvey et al. 2018; Kifle & Bekele, 2021; Woldegeorgis & Bekele, 2015) and data on monthly rainfall in the same locations (Fick & Hijmans, 2017; Harris et al. 2020; Jarvey et al. 2018). UF consumption was significantly associated with four tested rainfall estimates (Jarvey et al. 2018): same month total rainfall, $r_s = -0.43126$, $p(2-$

tailed) = .0003, monthly rainfall during the previous month, $r_s = -.349$, $p(2\text{-tailed}) = .00408$, cumulative rainfall during previous 2 months, $r_s = -.34862$, $p(2\text{-tailed}) = .00412$ and cumulative rainfall during previous 3 months, $r_s = -.2801$, $p(2\text{-tailed}) = .02274$ (Jarvey et al. 2018). Similar patterns are observed when the cumulative estimates include the month when UF consumption was considered (Figure S8A in Appendix 1). We verified if any region affected in a substantial way the results by repeating the analysis removing one dataset per time. The removal of the data from Indetu, but not of other datasets, substantially improves the correlation between monthly rainfall and UF usage (Figure S8B in Appendix 1).

The five datasets were collected in different years, but all included the same 6-month period (August–January). UF consumption during this period across the five locations is significantly different (One-Way ANOVA, $F_{4,25} = 3.82243$, $p = .014$), pairwise comparisons being significant for the Indetu–Sankaber pair (post-hoc Tukey test, $Q = 5.13$, $p = .01$; Figure S8C in Appendix 1). During this 6-month period, the average UF contribution to diet in Indetu (31.25%) was 50%–300% higher than in the other four regions (range 8.5%–22.21%). Using a moving window of seven consecutive months (the longest consecutive period available for Indetu), the distributions of estimated averages for Northern and Central overlap, but the mean value across seven continuous months at 33% in Southern is outside the variation reported for the other two taxa (Abu et al. 2018) (the maximum values in the four populations being 15%–26%; Figure S8C in Appendix 1; Abu et al. 2018). The main differences between Indetu and the other locations appear to be not in the maximum monthly UF consumption (46.6% in Indetu, 49.7% in Sankaber, 36.7% in Gich, 28.1% in Guassa, 57.7% in Wollo) but in this being consistently above 20% in Indetu (over the 9 months of available data) but not so in the other locations (3 months in Sankaber, 5 in Gich, 6 in Guassa and 7 in Wollo; Figure S8C in Appendix 1). We tested for differences in rainfall across the five regions but no significant differences were found for any of the four rainfall estimates related to the overlapping 6-month period (August–January) (one-way ANOVA, monthly rainfall during relevant month, $F_{4,25} = 0.03729$, $p = .99716$; monthly rainfall during the previous month, $F_{4,25} = 0.4303$, $p = .7854$; cumulative rainfall during previous 2 months, $F_{4,25} = 0.7754$, $p = .5516$; cumulative rainfall during previous 3 months, $F_{4,25} = 1.3515$, $p = .2789$). The same was also the case for cumulative estimates including the month when UF consumption was considered (data not shown).

3.4 - Discussion

3.4.1 - Population Dynamics and Climate Changes

The role played by climate variation on the evolutionary history of extant and extinct species has gathered centre stage in the discussion about major drivers of evolution (Carotenuto et al. 2016; Nogués-Bravo et al. 2010). Long and short-term wet-dry cycles have been investigated in relation to long-distance dispersals and isolation, events of gene-flow and hybridisation, drift and local adaptation in human and non-human primates (Chala et al. 2019; Stewart et al. 2020; Timmermann et al. 2022; Timmermann & Friedrich, 2016). Dietary and environmental specialisations (Chiou et al. 2022), together with a low fertility rate (one infant every 2–3 years; Dunbar, 1980) make gelada populations more vulnerable to eco-climatic changes, past and present. As such the reconstruction of how climate changed in the past can provide the context for interpreting the demographic events inferred from the analysis of genomic variation in gelada

populations. The analysis of autosomal microsatellites and mitochondrial DNA sequences in several *T. gelada* individuals sampled in the wild has previously suggested the subdivision of the species into three main populations, generally referred to as Northern, Central and Southern in accordance with their geographical distribution along Ethiopian highlands (Gippoliti, 2010; Shotake et al. 2016; Trede et al. 2020; Zinner et al. 2018). Our genomic analysis, extended to include data from gelada individuals collected from the Southern population, confirmed the proposed tripartition of this species. Full mitogenomes and Y chromosome sequences were similarly in agreement with the three-population structure, with further subdivisions hinted by mtDNA and genomic data, but not Y chromosome sequences (Caldon et al. 2024; Figure 2a,b in Appendix 1). Male lineages from different Northern subpopulations did not cluster according to their sampling location (unfortunately, having only one subpopulation for Central and Southern gelada, we were not able to explore this aspect in these populations) and lineages from different regions were related to each other. Such lack of structure might reflect the predominantly male-mediated intergroup migration observed in the wild (Dunbar, 1980; le Roux et al. 2011). The different demographic dynamics affecting males and females might also explain the significant differences in the amount of variation recovered using the mtDNA and the Y chromosome as summarised by their different TMRCAs, the Y chromosome being two times younger than the mitogenome (0.24 Mya vs. 0.49 Mya; Caldon et al. 2024). Similar differences in mitochondrial and Y chromosome TMRCAs have been noted previously in gorilla/orangutans but not in chimpanzees (Hallast et al. 2016). Such differences have been explained by non-mutually excluding scenarios related to the mating strategy implemented by males: (a) a substantial skew in male reproductive success, which would result in a reduction in Y chromosome diversity; (b) a reduced sperm competition, resulting in lower sperm production and an associated lower Y-related mutation rate. A lower mutation rate has not been reported in baboons, but male-skewed reproductive success has been described (Fischer et al. 2019; Wu et al. 2020). It is therefore possible that the skewed sampling process associated with male variation reproductive success might be the underlying explanatory mechanism for the observed Y/mtDNA TMRCAs in geladas. Relative differences in age are also dependent on the accuracy of mutation rates for different loci (Hallast et al. 2016). Further investigations into other species might provide useful insights on this topic.

The lack of shared Y chromosome and mtDNA lineages across the three gelada populations suggests a relatively deep subdivision among these groups, the mtDNA/Y chromosome TMRCAs between Southern and Central-Northern gelada dating to 500–150 kya (Caldon et al. 2024). Gene-flow among populations appears to have been interrupted only in the last 20 kya but population contractions had been in place much earlier than that (FIGURE 3.4). The lack of more recent coalescent events for uniparental markers could be due to sampling as well as drift, the latter expected to be more pronounced for haploid systems. The occurrence of deep mitochondrial and Y chromosome branches in Southern geladas is compatible with a model suggesting early colonisation of the region followed by periods of isolation and population contraction (Supplementary Material 1 in Caldon et al. 2024 online version; Figure S10 in Appendix 1) Interestingly, the most recent period of isolation suggested by genomic data appears to parallel changes in the climatic-driven habitat suitability of *T. gelada*, with gene-flow being interrupted since LGM (FIGURE 3.5; Figures 7 in Appendix 1). A dynamic, metapopulation model is suggested by admixture trees, PSMC plots and paleoclimate reconstructions, all depicting a scenario where events of expansion, gene-flow, isolation and possibly local extinction were probably not uncommon (Supplementary Material 1 in Caldon et al. 2024 online version).

The extreme fluctuations in habitat suitability are also compatible with the continuous reduction in effective population size estimated for *T. gelada* (Figure S10 in Appendix 1; Supplementary Material 1 in Caldon et al. 2024 online version; FIGURE 3.5), resulting from decreasing population size and/or connectivity. The more extreme reduction in diversity and increase in RoH reported for Southern geladas parallel the more severe change in effective population size highlighted over recent times by PSMC and follows the lower degree of habitat suitability experienced over time by the Southern population (FIGURE 3.3-3.5; Figure 7 in Appendix 1; Figure S10 in Appendix 1; Supplementary Material 1 in Caldon et al. 2024 online version).

3.4.2 - Hybridisation

Our extensive genomic comparison provided support for previously suggested gene-flow between *Theropithecus* and *Papio*, their last common ancestor possibly dating to 5 Ma, close to the human–chimpanzee split (Liedigk et al. 2014). Hybrids have been potentially identified in the wild (Dunbar & Dunbar, 1974) and observed in captivity (Jolly et al. 1997; Markarjan et al. 1974). The phenotypes of the captive and wild hybrids were different, possibly reflecting gelada or *Papio* backcrossing (Jolly et al. 1997). Captive F1 showed intermediate or more gelada-like features, exceeding both parental populations for body size (Jolly et al. 1997; Markarjan et al. 1974). Our data suggested ancient gelada contributions to *Papio* species present in Ethiopia (*P. hamadryas* and *P. anubis*). In particular the highest signal of hybridisation is between the Southern gelada population and the Ethiopian *P. anubis* population. Interestingly, the time of the last contact of *Papio anubis* with Southern geladas aligned with a period when the predicted Southern geladas suitable habitat area was increasing (Figure S10 in Appendix 1; Supplementary Material 1 in Caldon et al. 2024 online version). Similar changes have occurred multiple times in the past and might have facilitated also the more ancient contacts of *T. gelada* with *P. hamadryas*.

The occurrence of captive fertile F1 (Markarjan et al. 1974) but the lack of genomic evidence for recent gene-flow could be explained by the hybrid F1 being generally disfavoured in the wild and therefore less likely to survive and reproduce. Jolly and colleagues (Jolly et al. 1997) proposed two possible explanations of why hybrids might have a lower fitness in the wild, both based on their larger appearance. Both considered the additional energy requirements associated with larger body size: increased nutritional requirements but lower-crowned teeth (as observed in the hybrids in the wild; Jolly et al. 1997) might drive early loss of dental function and inability to fulfil their nutritional needs, and the increased maternal burden associated with the raising of a larger foetus and the nursing of a bigger infant (Jolly et al. 1997).

The antiquity of the introgression events probably provided enough time for natural selection to remove incompatible or deleterious gelada alleles from the genome of *Papio* (and *Lophocebus*), as well as increase the frequency of favourable variants, in a scenario not too dissimilar from what was observed in humans after the hybridisation with Neanderthals and Denisovans (Tobler et al. 2023). The recent release of a large *Papio* genomic dataset is expected to enable a more detailed investigation of the genomic distribution of gelada ancestry and their possible phenotypic impact (Groh & Coop, 2024; Sørensen et al. 2023).

3.4.3 - Climate, Diet and Local Adaptations

The unique dietary profile of geladas dominated by grass makes the variation across populations in UF usage of interest (Figure S8 in Appendix 1; Jarvey et al. 2018). Grass is the staple of choice of geladas, its contribution to diet being shaped by its availability ultimately influenced by rainfall. UF operates as fallback food, its usage strongly driven by grass availability. While the link between grass consumption and rainfall is strong, the one between dietary UF and rainfall is present but weaker (Jarvey et al. 2018 and our results). Other factors probably play a role in determining UF usage (Abu et al. 2018). All this considered, despite no variation in rainfall, geladas of the Southern population of Indetu appear to consume more UF than other gelada groups. Interestingly Southern geladas harbour more copies of LOC112607926, one of the two genes annotated as pancreatic amylase in the gelada reference genome (FIGURE 3.6), a genomic feature observed in dogs, pigs and human populations with a diet enriched in starchy food (Jarvey et al. 2018; Pajic et al. 2019; Perry et al. 2007). The reduction in diversity around the duplicated locus in Southern gelada is indicative of positive selection, while the chromosomal extent of this pattern suggests this being a relatively recent event. For comparison, the reduction of diversity around the variants associated with lactase persistence in humans is similarly observed over more than 1 Mb (Bersaglieri et al. 2004; Joslin et al. 2020). Considering that selection on the persistence of LCT expression into adulthood has been dated to no more than a few thousand years ago (considering a generation time in humans of 28 years, these correspond to a few hundred generations; Burger et al. 2020; Evershed et al. 2022), it is reasonable to consider the pattern observed in Southern gelada being generated over a number of generations around the same order of magnitude (at least hundreds).

During the past 30 kya, Southern Ethiopia has undergone strong climatic changes, from dry and relatively cold during the Last Glacial Maximum (LGM, 25–18 kya) to the African Humid Period (AHP, 15–5 kya; Casas-Gallego et al. 2023; Fischer et al. 2021; Foerster et al. 2012). Our past SDM predictions perfectly fit the Ethiopian climatic trends suggesting that gelada spatial distribution was affected by these intense climatic fluctuations (Figure 7 in Appendix 1). Particularly, we show that during the LGM the Southern geladas extremely suffered cold and dry conditions (Figure 7 in Appendix 1; Supplementary Material 1 in Caldon et al. 2024 online version; Figure S10 in Appendix 1) whereas the impact was limited for the other two populations. We propose that increased UF consumption and amylase copy number are related in Southern geladas. The ecological context experienced over time by this population might have driven positive selection on the amylase duplication in the last few thousand years, compatible with the physical extent of the reduction of diversity around this locus. Increased UF consumption by geladas has been observed when grass/rain is scarce and metabolic adaptation enabling more efficient processing of starchy food is expected to provide an evolutionary advantage in such situations (Jarvey et al. 2018; Pajic et al. 2019; Perry et al. 2007). While past climate might have provided the context for positively selecting a higher number of copies of the amylase gene in Southern geladas, other dynamics might have contributed over time too. The observation that increased UF consumption in Indetu occurred despite no differences in rainfall might reflect behavioural differences in food preference across gelada populations. It has been also reported that gelada UF consumption in Indetu in the wet season is slightly higher than during the dry season (seasonal mean 38% and 33% respectively; Abu et al. 2018). A general, behavioural-related, increase in UF consumption could have reinforced the climate-related selective pressure on salivary amylase copy-number, generating a co-evolutionary scenario not too

dissimilar from the gene–culture interactions reported in humans for lactase persistence (Evershed et al. 2022). It remains to be tested if in Indetu the use of UF is less related to the amount of grass available and if local behavioural traits contribute to the increase in UF consumption (Whiten et al. 1999). Interestingly, the removal of data from Indetu improves the correlation between UF consumption and rainfall in the dataset we tested, suggesting that other variables, including anthropogenic, affect dietary UF contribution to diet in Indetu (Abu et al. 2018).

The Ethiopian Highlands (part of the Eastern Afromontane Biodiversity Hotspot; Mittermeier et al. 2011) can be divided into the main North-Western massif (Western plateau) and the smaller South-Eastern plateau, the two separated by the East African Great Rift Valley (Yalden et al. 1996). The highland plateaus belong to the same ecozone (Afroalpine and subafroalpine vegetation) although, some variation in the taxonomic composition of both plant and animal communities as a result of biogeographical patterns across gelada locations is present (Vivero et al. 2006). Currently, the lack of a systematic investigation in plant composition across the plateaus and locations of gelada occurrence prevents any further analysis in how these affect food consumption. Nevertheless, it is worth noticing that gelada on the Eastern plateaus live at slightly lower altitudes than those on the Western plateau (1600–2300 m vs. above 2300; Abu et al. 2018; Belay & Shotake, 1998), which might further affect the set of species available to geladas. However, it will be necessary to collect more information on what plant species are present and the extent of their occurrence across sites to test how these aspects shape what geladas prefer and ultimately consume.

We note that the two Southern Individuals here whole genome sequenced were not collected in Indetu and it remains to be shown if increased copy number/increased UF consumption are features shared across all Southern geladas. More ecological and behavioural data, as well as a more extensive characterisation of amylase copy-number variation and their expression in Southern geladas, are necessary to test the association between UF consumption, grass availability and amylase genes in this population.

We note that an increase in the copy number of pancreatic amylase is expected to increase not only the expression in the pancreas but also in the salivary glands, as pancreatic amylase is also expressed in salivary glands (Axelsson et al. 2013; Fagerberg et al. 2014; Han et al. 2022; Mau et al. 2010; Pajic et al. 2019; Samuelson et al. 1996; Torres et al. 2018). Selective pressure associated with an increase in starch consumption have been shown to favour the expansion of amylase genes, salivary and/or pancreatic (Pajic et al. 2019). The related increase in amylase production, both in the pancreas and the salivary glands, results in an increased capacity of processing starch both in the mouth and the digestive system (Axelsson et al. 2013; Mau et al. 2010; Pajic et al. 2019; Perry et al. 2007; Samuelson et al. 1996), which might be the case also for Southern geladas.

3.4.4 - Implications for the Taxonomy and Conservation of *T. gelada*

The genomic analyses here presented support that presently the genus *Theropithecus* may be subdivided into three evolutionary units that have been taxonomically designated as *Theropithecus gelada gelada* (Rüppell, 1835), *Theropithecus gelada obscurus* Heuglin, 1863 respectively in the Northern and central areas of the Ethiopian plateau, and, in the South-Eastern mountain range, *Theropithecus gelada arsi*, so agreeing with most recent literature (Gippoliti,

2010; Shotake et al. 2016; Trede et al. 2020; Zinner et al. 2018). This taxonomic account represents a solid base for conservation policies that aim to protect the current diversity of the genus, but, as often the case, must not be considered definitive. In particular, it remains to be determined whether the three identified evolutionary units should be treated as having equal rank or if the Arsi (Southern) gelada should be considered a distinct species. This consideration is based on the potentially older divergence of the Arsi gelada from the two subspecies, gelada (Northern) and obscurus (Central), suggested by deep coalescent events of uniparental markers and is consistent with the Rift Valley's well-known role as a barrier for mountain animal specialists in Ethiopia (Lavrenchenko & Bekele, 2017). Additionally, the ecological challenges of the Arsi habitat, supported by food-related adaptations reported for the amylase gene and the specific present and past climatic niche of the Arsi geladas, further support treating the Arsi population as a distinct species. We highlight that other standard taxonomic databases, such as cranial morphology, have been scarcely investigated so far and could be critical to further refine our taxonomic knowledge.

In this regard, we also note that the northernmost gelada populations have been recently confirmed in the Tigray, whose genetic and morphological affinity to other gelada populations is still unclear (Girmay & Dati, 2020; Girmay & Tesfay, 2020; Haileselasie et al. 2023). Considering the geographical features of the region, it is not impossible that the geladas from Tigray will turn out to be closer to *T. g. obscurus* (Central) than *T. g. gelada* (Northern), making the geographical designation of these subspecies as Central and Northern incorrect and instead supporting an alternative nomenclature as Western (gelada) and Eastern (obscurus) geladas (Gippoliti, 2010). Future characterisation of these groups will clarify this interesting geographical and taxonomic aspect of the species.

As a species *Theropithecus gelada* is considered Least Concern by IUCN (Gippoliti et al. 2019). Irrespective of the taxonomic rank accorded, *T. gelada arsi*, with between 2000 and 3000 surviving individuals estimated (De Jong et al. 2023), is a conservation priority despite the lack of a formal description that has delayed a formal recognition by IUCN (Gippoliti, 2022). The Arsi gelada may be considered Endangered according to IUCN criteria – a formal assessment has just been published (De Jong et al. 2023). The Galama Ridge, which offers protection to the Arsi geladas, suffers from intense anthropogenic pressure and no protected areas exist in the region. The remaining mosaic forests are threatened by intensive deforestation and agricultural ploughing. Montane habitats suffer from burning during the dry season and overgrazing during the wet season (Kostin et al. 2019). This taxon may therefore serve as a much-needed flagship species attracting naturalistic tourism to their narrow region of occupancy along the gorges of the Webi Shebeli and Robi rivers. Obviously, revenue from tourism must be shared with local communities to lessen anthropogenic impacts on ecosystems.

3.5 - Conclusions

By integrating genomic, climatic and dietary data we have generated a detailed picture of the evolutionary history of *Theropithecus gelada*. The inclusion of more genomic data, from living individuals and museum specimens (Gippoliti, 2010), is expected to refine our results by providing a more detailed description of the degree of population structure and gene flow across

populations that may be critical for long-term conservation planning. The integration of morphological data and the inclusion of functional analyses will contribute additional insights on the degree of variation within the species and the evolutionary significance of such diversity. Reconstructing the history of this species is also expected to unlock to what extent geladas will be able to face the challenges related to ongoing and future climatic changes (Ahmed et al. 2023). Indeed, a formal assessment of this population has just been published (De Jong et al. 2023) that suggests the Arsi gelada should be classified as Endangered according to IUCN criteria.

3.6 - References

Abu, K., Mekonnen, A., Bekele, A., & Fashing, P. J. (2018). Diet and activity patterns of Arsi geladas in low-elevation disturbed habitat South of the Rift Valley at Indetu, Ethiopia. *Primates; journal of primatology*, 59(2), 153–161. <https://doi.org/10.1007/s10329-017-0640-9>

Ahmed, A. S., Chala, D., Kufa, C. A., Atickem, A., Bekele, A., Svenning, J. C., & Zinner, D. (2023). Potential changes in the extent of suitable habitats for geladas (*Theropithecus gelada*) in the Anthropocene. *BMC ecology and evolution*, 23(1), 65. <https://doi.org/10.1186/s12862-023-02173-3>

Almarri, M. A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A. S., Chen, Y., Hurles, M. E., Tyler-Smith, C., & Xue, Y. (2020). Population Structure, Stratification, and Introgression of Human Structural Variation. *Cell*, 182(1), 189–199.e15. <https://doi.org/10.1016/j.cell.2020.05.024>

Axelsson, E., Ratnakumar, A., Arendt, M. L., Maqbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, A., & Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, 495(7441), 360–364. <https://doi.org/10.1038/nature11837>

Batra, S. S., Levy-Sakin, M., Robinson, J., Guillory, J., Durinck, S., Vilgalys, T. P., Kwok, P. Y., Cox, L. A., Seshagiri, S., Song, Y. S., & Wall, J. D. (2020). Accurate assembly of the olive baboon (*Papio anubis*) genome using long-read and Hi-C data. *GigaScience*, 9(12), g1a134. <https://doi.org/10.1093/gigascience/giaa134>

de Beaux, O. (1925). Mammiferi dell'Abissinia raccolti dal Signor Ugo Ignesti. *Atti Soc Ital Sci Nat Milano* 64: 196–218.

Belay, G., Shotake, T. Blood protein variation of a new population of gelada baboons (*Theopithecus gelada*), in the Southern Rift Valley, Arsi Region, Ethiopia. *Primates* 39, 183–198 (1998). <https://doi.org/10.1007/BF02557730>

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American journal of human genetics*, 74(6), 1111–1120.

<https://doi.org/10.1086/421051>

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappel, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., & Barillot, E. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* (Oxford, England), 28(3), 423–425. <https://doi.org/10.1093/bioinformatics/btr670>

Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y., Žegarac, A., Hofmanová, Z., Winkelbach, L., Reyna-Blanco, C. S., Bieker, V., Orschiedt, J., Brinker, U., Scheu, A., Leuenberger, C., Bertino, T. S., Bollongino, R., Lidke, G., Stefanović, S., ... Wegmann, D. (2020). Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 3,000 Years. *Current biology: CB*, 30(21), 4307–4315.e13. <https://doi.org/10.1016/j.cub.2020.08.033>

Cahill, J. A., Soares, A. E., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1699), 20150138. <https://doi.org/10.1098/rstb.2015.0138>

Carotenuto, F., Di Febbraro, M., Melchionna, M., Castiglione, S., Saggese, F., Serio, C., Mondanaro, A., Passaro, F., Loy, A., Raia, P. (2016). The influence of climate on species distribution over time and space during the late Quaternary. *Quaternary Science Reviews* 149, 188–199. <https://doi.org/10.1016/j.quascirev.2016.07.036>.

Casas-Gallego, M., Hahn, K., Neumann, K., Demissew, S., Schmidt, M., Bodin, S. C., & Bruch, A. A. (2023). Publisher Correction: Cooling-induced expansions of Afromontane forests in the Horn of Africa since the Last Glacial Maximum. *Scientific reports*, 13(1), 10687. <https://doi.org/10.1038/s41598-023-37919-y>

Chala, D., Roos, C., Svenning, J. C., & Zinner, D. (2019). Species-specific effects of climate change on the distribution of suitable baboon habitats - Ecological niche modeling of current and Last Glacial Maximum conditions. *Journal of human evolution*, 132, 215–226. <https://doi.org/10.1016/j.jhevol.2019.05.003>

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* (Oxford, England), 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>

Chiou, K. L., Janiak, M. C., Schneider-Crease, I. A., Sen, S., Ayele, F., Chuma, I. S., Knauf, S., Lemma, A., Signore, A. V., D'Ippolito, A. M., Abebe, B., Haile, A. A., Kebede, F., Fashing, P. J., Nguyen, N., McCann, C., Houck, M. L., Wall, J. D., Burrell, A. S., Bergey, C. M., ... Snyder-Mackler, N. (2022). Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. *Nature ecology & evolution*, 6(5), 630–643.

<https://doi.org/10.1038/s41559-022-01703-4>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics* (Oxford, England), 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>

De Jong, Y.A., Moges, A., Jenner, T., et al. *Theropithecus gelada* ssp. *Arsi*. The IUCN Red List of Threatened Species 2023 [provisional].

Delson, E. (1993). The fossils from Africa and India and the taxonomy of the genus, In Jablonski, N. G. (ed.), *Theropithecus: Rise and Fall of a Primate Genus*. Cambridge University Press, Cambridge, pp. 157–190.

Dunbar, R.I.M., Dunbar, P. (1974). On hybridization between *Theropithecus gelada* and *Papio anubis* in the wild. *Journal of Human Evolution* 3, 3. [https://doi.org/10.1016/0047-2484\(74\)90176-6](https://doi.org/10.1016/0047-2484(74)90176-6).

Dunbar, R.I.M. (1977). Feeding ecology of gelada baboons: a preliminary report. In: T.H. Clutton-Brock ed. *Primate ecology: studies of feeding and ranging behaviour in lemurs, monkeys and apes*. Academic Press, London: 251–273.

Dunbar, R.I.M. (1980). Demographic and Life History Variables of a Population of Gelada Baboons (*Theropithecus gelada*). *Journal of Animal Ecology*, 49 (2), 485–506.

Dunbar, R.I.M. (1993). Socioecology of the extinct theropithecids: a modelling approach. , In Jablonski, N. G. (ed.), *Theropithecus: Rise and Fall of a Primate Genus*. Cambridge University Press, Cambridge, pp. 157–190.

Evershed, R. P., Davey Smith, G., Roffet-Salque, M., Timpson, A., Diekmann, Y., Lyon, M. S., Cramp, L. J. E., Casanova, E., Smyth, J., Whelton, H. L., Dunne, J., Brychova, V., Šoberl, L., Gerbault, P., Gillis, R. E., Heyd, V., Johnson, E., Kendall, I., Manning, K., Marciniak, A., ... Thomas, M. G. (2022). Dairying, diseases and the evolution of lactase persistence in Europe. *Nature*, 608(7922), 336–345. <https://doi.org/10.1038/s41586-022-05010-7>

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpour, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigartyo, C. A., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., ... Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & cellular proteomics: MCP*, 13(2), 397–406. <https://doi.org/10.1074/mcp.M113.035600>

Fashing, P. J., Nguyen, N., Venkataraman, V. V., & Kerby, J. T. (2014). Gelada feeding ecology in an intact ecosystem at Guassa, Ethiopia: Variability over time and implications for theropithecids and hominin dietary evolution. *American journal of physical anthropology*, 155(1), 1–16. <https://doi.org/10.1002/ajpa.22559>

Fick, S.E., Hijmans, R.J. (2017). WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *International Journal of Climatology* 37, 12.

Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., Seyfarth, R. M., ... Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. *eLife*, 8, e50989. <https://doi.org/10.7554/eLife.50989>

Fischer, M.L., Bachofer, F., Yost, C.L., Bludau, I.J.E., Schepers, C., Foerster, V., Lamb, H., Schäbitz, F., Asrat, A., Trauth, M.H., et al. (2021). A Phytolith Supported Biosphere-Hydrosphere Predictive Model for Southern Ethiopia: Insights into Paleoenvironmental Changes and Human Landscape Preferences since the Last Glacial Maximum. *Geosciences*, 11, 418. <https://doi.org/10.3390/geosciences11100418>

Foerster, V., Junginger, A., Langkamp, O., Gebru, T., Asrat, A., Umer, M., Lamb, H. F., Wennrich, V., Rethemeyer, J., Nowaczyk, N., Trauth, M. H., & Schaebitz, F. (2012). Climatic change recorded in the sediments of the Chew Bahir basin, Southern Ethiopia, during the last 45,000 years. *Quaternary International*, 274, 25–37. <https://doi.org/10.1016/j.quaint.2012.06.028>

Gippoliti S. (2010). *Theropithecus gelada* distribution and variations related to taxonomy: history, challenges and implications for conservation. *Primates; journal of primatology*, 51(4), 291–297. <https://doi.org/10.1007/s10329-010-0202-x>

Gippoliti, S., Mekonnen, A., Burke, R., Nguyen, N. & Fashing, P.J. (2019). *Theropithecus gelada*. The IUCN Red List of Threatened Species.

Gippoliti, S. (2020). Everything mammal conservation biologists always wanted to know about taxonomy (but were afraid to ask). *Journal for Nature Conservation* 54, 125793. <https://doi.org/10.1016/j.jnc.2020.125793>.

Gippoliti, S. (2022). Taxonomic history affects current conservation efforts: examples from the primates of the Horn of Africa. *Journal of Animal Diversity*, 4 (1): 18 –20. <http://dx.doi.org/10.52547/JAD.202.2.4.1.2>

Girmay, T., Dati, D. (2020). Population size, group and age structure of geladas (*Theropithecus gelada*) in escarpments of Eastern Tigray, Ethiopia: implication for conservation. *j ecology environ* 44, 20. <https://doi.org/10.1186/s41610-020-00163-w>

Girmay, T., & Tesfay, H. (2020). New geographical range of geladas (*Theropithecus gelada*) in Tigray region, Northern Ethiopia. *African Journal of Ecology*, 59(1), 277–280. <https://doi.org/10.1111/aje.12818>

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H. J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., Auvil, L., ... Schook, L.

B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491(7424), 393–398. <https://doi.org/10.1038/nature11622>

Groh, J. S., & Coop, G. (2024). The temporal and genomic scale of selection following hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 121(12), e2309168121. <https://doi.org/10.1073/pnas.2309168121>

Groves, C.P., Cotterill, F.P.D., Gippoliti, S. et al. Species definitions and conservation: a review and case studies from African mammals. *Conserv Genet* 18, 1247–1256 (2017). <https://doi.org/10.1007/s10592-017-0976-0>

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS genetics*, 15(7), e1008302. <https://doi.org/10.1371/journal.pgen.1008302>

Haileselasie, T. H., Muluallem, G., & Welegerima, K. (2023). The first large-scale record of gelada's spatial distribution and population size in Mountain chains of Tigray, Northern Ethiopia. *Heliyon*, 9(9), e19346. <https://doi.org/10.1016/j.heliyon.2023.e19346>

Hallast, P., Maisano Delser, P., Batini, C., Zadik, D., Rocchi, M., Schempp, W., Tyler-Smith, C., & Jobling, M. A. (2016). Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. *Genome research*, 26(4), 427–439. <https://doi.org/10.1101/gr.198754.115>

Han, L., Wei, X., Liu, C., Volpe, G., Zhuang, Z., Zou, X., Wang, Z., Pan, T., Yuan, Y., Zhang, X., Fan, P., Guo, P., Lai, Y., Lei, Y., Liu, X., Yu, F., Shangguan, S., Lai, G., Deng, Q., Liu, Y., ... Liu, L. (2022). Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. *Nature*, 604(7907), 723–731. <https://doi.org/10.1038/s41586-022-04587-3>

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data*, 7(1), 109. <https://doi.org/10.1038/s41597-020-0453-3>

Heuglin T., von (1863) Beiträge zur Zoologie Afrikas. Über einige Säugethiere des Bäschlo-Gebietes. *Nova Acta Academiae Caesarea Leopoldio, Carolinae Naturae*, 30: 1–14.

Iwamoto T. (1993a). The ecology of *Theropithecus gelada*. In Jablonski, N. G. (ed.), *Theropithecus: Rise and Fall of a Primate Genus*. Cambridge University Press, Cambridge, pp. 157–190.

Iwamoto T. (1993b). Food digestion and energetic conditions in *Theropithecus gelada*. In Jablonski, N. G. (ed.), *Theropithecus: Rise and Fall of a Primate Genus*. Cambridge University Press, Cambridge, pp. 157–190.

Jarvey, J. C., Low, B. S., Pappano, D. J., Bergman, T. J., Beehner, J. C. (2018). Graminivory and Fallback Foods: Annual Diet Profile of Geladas (*Theropithecus Gelada*) Living in the Simien Mountains National Park, Ethiopia. *International Journal of Primatology* 39, 1.

- Jolly, C.J., Woolley-Barker, T., Beyene, S. et al. (1997). Intergeneric Hybrid Baboons. *International Journal of Primatology* 18, 597–627. <https://doi.org/10.1023/A:1026367307470>
- Joslin, S. E. K., Durbin-Johnson, B. P., Britton, M., Settles, M. L., Korf, I., & Lemay, D. G. (2020). Association of the Lactase Persistence Haplotype Block With Disease Risk in Populations of European Descent. *Frontiers in genetics*, 11, 558762. <https://doi.org/10.3389/fgene.2020.558762>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Keppel, G., Van Niel, K. P., Wardell-Johnson, G. W., Yates, C. J., Byrne, M., Mucina, L., ... & Franklin, S. E. (2012). Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global ecology and biogeography*, 21(4), 393-404.
- Kifle, Z., & Bekele, A. (2021). Feeding ecology and diet of the Southern geladas (*Theropithecus gelada obscurus*) in human-modified landscape, Wollo, Ethiopia. *Ecology and evolution*, 11(16), 11373–11386. <https://doi.org/10.1002/ece3.7927>
- King, M. (1995). Species evolution: the role of chromosome change. Cambridge University Press 44, 4, 578–580. <https://doi.org/10.1093/sysbio/44.4.578>.
- Kostin, D., Kasso, M., Komarova, V., Martynov, A., Gromov, A., Alexandrov, D., Bekele, A., Zewdie, C., Bryja, J. & Lavrenchenko, L. (2019). Taxonomic and genetic diversity of rodents from the Arsi Mountains (Ethiopia). *Mammalia*, 83(3), 237-247. <https://doi.org/10.1515/mammalia-2017-0135>
- Lavrenchenko, L.A., Bekele, A. (2017). Diversity and conservation of Ethiopian mammals: what have we learned In 30 Years? *Ethiopian Journal of Biological Sciences* 16,1.
- le Roux, A., Beehner, J. C., & Bergman, T. J. (2011). Female philopatry and dominance patterns in wild geladas. *American journal of primatology*, 73(5), 422–430. <https://doi.org/10.1002/ajp.20916>
- Liedigk, R., Roos, C., Brameier, M., & Zinner, D. (2014). Mitogenomics of the Old World monkey tribe Papionini. *BMC evolutionary biology*, 14, 176. <https://doi.org/10.1186/s12862-014-0176-1>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

- Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* (Oxford, England), 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., & Reich, D. (2023). On the limits of fitting complex models of population history to f-statistics. *eLife*, 12, e85492. <https://doi.org/10.7554/eLife.85492>
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* (Oxford, England), 26(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
- Markarjan, D.S., Isakov, E.P., Kondakov, G.I. (1974). Intergeneric hybrids of the lower (42-chromosome) monkey species of the Sukhumi monkey colony. *Journal of Human Evolution* 3, 3. [https://doi.org/10.1016/0047-2484\(74\)90183-3](https://doi.org/10.1016/0047-2484(74)90183-3).
- Mau, M., Südekum, K. H., Johann, A., Sliwa, A., & Kaiser, T. M. (2010). Indication of higher salivary alpha-amylase expression in hamadryas baboons and geladas compared to chimpanzees and humans. *Journal of medical primatology*, 39(3), 187–190. <https://doi.org/10.1111/j.1600-0684.2010.00407.x>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., Gascon, C. (2011). Global Biodiversity Conservation: The Critical Role of Hotspots. In: Zachos, F., Habel, J. (eds) *Biodiversity Hotspots*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20992-5_1
- Montinaro, F., & Capelli, C. (2020). A Worldwide Map of Human Structural Variants. *Trends in genetics: TIG*, 36(10), 722–725. <https://doi.org/10.1016/j.tig.2020.07.002>
- Mori, A., & Belay, G. (1990). The distribution of baboon species and a new population of gelada baboons along the Wabi-Shebeli river, Ethiopia. *Primates*, 31, 495-508.
- Nogués-Bravo, D., Ohlemüller, R., Batra, P., & Araújo, M. B. (2010). Climate predictors of late quaternary extinctions. *Evolution; international journal of organic evolution*, 64(8), 2442–2449. <https://doi.org/10.1111/j.1558-5646.2010.01009.x>
- Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R. A., Garneau, D., Daugherty, E., Globig, A., Ruhl, S., & Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *eLife*, 8, e44628. <https://doi.org/10.7554/eLife.44628>

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., Kessing, B., Pontius, J., Roelke, M., Rumppler, Y., Schneider, M. P., Silva, A., O'Brien, S. J., & Pecon-Slaterry, J. (2011). A molecular phylogeny of living primates. *PLoS genetics*, 7(3), e1001342. <https://doi.org/10.1371/journal.pgen.1001342>

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature genetics*, 39(10), 1256–1260. <https://doi.org/10.1038/ng2123>

Pigkford, M. (1993). Climate change, biogeography and *Theropithecus*. In: Jablonski NG, ed. *Theropithecus: The Rise and Fall of a Primate Genus*. Cambridge: Cambridge University Press: 227–244.

Raia, P., Passaro, F., Fulgione, D., & Carotenuto, F. (2012). Habitat tracking, stasis and survival in Neogene large mammals. *Biology letters*, 8(1), 64–66. <https://doi.org/10.1098/rsbl.2011.0613>

Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American journal of physical anthropology*, 174(3), 407–417. <https://doi.org/10.1002/ajpa.24186>

Rüppell W.P.E.S. (1835) *Neue Wirbelthiere zu der Fauna von Abbyssinien gehörig. Säugethiere*. Siegmund Schmerber, Frankfrut-am-Main.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909), 832–837. <https://doi.org/10.1038/nature01140>

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., ... Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. <https://doi.org/10.1038/nature06250>

Samuelson, L. C., Phillips, R. S., & Swanberg, L. J. (1996). Amylase gene structures in primates: retroposon insertions and promoter evolution. *Molecular biology and evolution*, 13(6), 767–779. <https://doi.org/10.1093/oxfordjournals.molbev.a025637>

Santander, C., Molinaro, L., Mutti, G., Martínez, F. I., Mathe, J., Ferreira da Silva, M. J., Caldon, M., Oteo-Garcia, G., Aldeias, V., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Hammond, P., Lüdecke, T., Pinto, M. J., Meira Paulo, L., Stalmans, M., Regala, F. T., ... Capelli, C. (2022). Genomic variation in baboons from central Mozambique unveils complex evolutionary relationships with other *Papio* species. *BMC ecology and evolution*, 22(1), 44. <https://doi.org/10.1186/s12862-022-01999-7>

Shotake, T., Saijuntha, W., Agatsuma, T. et al. (2016). Genetic diversity within and among gelada (*Theropithecus gelada*) populations based on mitochondrial DNA analysis, *Anthropological Science* 124, 3. DOI: 10.1537/ase.160717

Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., Storer, J. M., Kuhlwilm, M., Fonsere, C., Seshadri, L., Bergey, C. M., Burrell, A. S., Bergman, J., Phillips-Conroy, J. E., Shiferaw, F., Chiou, K. L., Chuma, I. S., Keyyu, J. D., Fischer, J., Gingras, M. C., ... Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science (New York, N.Y.)*, 380(6648), eabn8153. <https://doi.org/10.1126/science.abn8153>

Stewart, B.M., Turner, S.E., Matthews H.D. (2020). Climate Change Impacts on Potential Future Ranges of Non-Human Primate Species. *Climatic Change* 162, 4.

Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B. P., Baker, C., Nordenfelt, S., Bamshad, M., Jorde, L. B., Posukh, O. L., Sahakyan, H., Watkins, W. S., Yepiskoposyan, L., Abdullah, M. S., Bravi, C. M., Capelli, C., Hervig, T., Wee, J. T., ... Eichler, E. E. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science (New York, N.Y.)*, 349(6253), aab3761. <https://doi.org/10.1126/science.aab3761>.

Timmermann, A., & Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. *Nature*, 538(7623), 92–95. <https://doi.org/10.1038/nature19365>

Timmermann, A., Yun, K. S., Raia, P., Ruan, J., Mondanaro, A., Zeller, E., Zollikofer, C., Ponce de León, M., Lemmon, D., Willeit, M., & Ganopolski, A. (2022). Climate effects on archaic human habitats and species successions. *Nature*, 604(7906), 495–501. <https://doi.org/10.1038/s41586-022-04600-9>

Tobler, R., Souilmi, Y., Huber, C. D., Bean, N., Turney, C. S. M., Grey, S. T., & Cooper, A. (2023). The role of genetic selection and climatic factors in the dispersal of anatomically modern humans out of Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 120(22), e2213061120. <https://doi.org/10.1073/pnas.2213061120>

Torres, S. M. F., Furrow, E., Souza, C. P., Granick, J. L., de Jong, E. P., Griffin, T. J., & Wang, X. (2018). Salivary proteomics of healthy dogs: An in depth catalog. *PloS one*, 13(1), e0191307.

<https://doi.org/10.1371/journal.pone.0191307>

Trauth, M. H., Deino, A. L., Bergner, A. G., & Strecker, M. R. (2003). East African climate change and orbital forcing during the last 175 kyr BP. *Earth and Planetary Science Letters*, 206(3-4), 297-313.

Trede, F., Lemkul, A., Atickem, A., Beehner, J.C., Bergman, T.J., Burke, R.J., Fashing, P.J., Knauf, S., Mekonnen, A., Moges, A., Nguyen, N., Roos, C., & Zinner, D. (2020). Geographic distribution of microsatellite alleles in geladas (Primates, Cercopithecidae): Evidence for three evolutionary units. *Zoologica Scripta*, 49, 659 - 667.

Vasimuddin, M., Misra, S., Li, H., Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, pp. 314-324.

Vivero, J.L., Ensermu, K. & Sebsebe, D. (2006). Progress on the Red List of plants of Ethiopia and Eritrea: conservation and biogeography of endemic flowering taxa. In: S.A. Ghazanfar & H.J. Beentje (eds), *Taxonomy and ecology of African plants, their conservation and sustainable use*, pp. 761–778. Royal Botanic Gardens, Kew.

Walker, J. A., Jordan, V. E., Storer, J. M., Steely, C. J., Gonzalez-Quiroga, P., Beckstrom, T. O., Rewerts, L. C., St Romain, C. P., Rockwell, C. E., Rogers, J., Jolly, C. J., Konkel, M. K., Baboon Genome Analysis Consortium, & Batzer, M. A. (2019). Alu insertion polymorphisms shared by *Papio baboons* and *Theropithecus gelada* reveal an intertwined common ancestry. *Mobile DNA*, 10, 46. <https://doi.org/10.1186/s13100-019-0187-y>

Warren, W. C., Harris, R. A., Haukness, M., Fiddes, I. T., Murali, S. C., Fernandes, J., Dishuck, P. C., Storer, J. M., Raveendran, M., Hillier, L. W., Porubsky, D., Mao, Y., Gordon, D., Vollger, M. R., Lewis, A. P., Munson, K. M., DeVogelaere, E., Armstrong, J., Diekhans, M., Walker, J. A., ... Eichler, E. E. (2020). Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science (New York, N.Y.)*, 370(6523), eabc6617. <https://doi.org/10.1126/science.abc6617>

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, 399(6737), 682–685. <https://doi.org/10.1038/21415>

Wilson Sayres, M. A., Venditti, C., Pagel, M., & Makova, K. D. (2011). Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution; international journal of organic evolution*, 65(10), 2800–2815. <https://doi.org/10.1111/j.1558-5646.2011.01337.x>

Woldegeorgis, C., Bekele A. (2015). Activity budget and behavioural patterns of Gelada *Theropithecus gelada* (Mammalia: Primates: Cercopithecidae) on the Gich Plateau of the Simien Mountains National Park, Ethiopia. *Journal of Threatened Taxa* 7, 8.

- Wu, F. L., Strand, A. I., Cox, L. A., Ober, C., Wall, J. D., Moorjani, P., & Przeworski, M. (2020). A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS biology*, 18(8), e3000838. <https://doi.org/10.1371/journal.pbio.3000838>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, 2(3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Yalden, D. W., Lagen, M. J., Kock, D., & Hillman, J. C. (1996). Catalogue of the mammals of Ethiopia and Eritrea. 7. Revised checklist, zoogeography and conservation. *Tropical Zoology*, 9(1), 73–164. <https://doi.org/10.1080/03946975.1996.10539304>
- Yalden, D.W., Lagen, M.J., Kock, D. CATALOGUE OF THE MAMMALS OF ETHIOPIA. *Monitore Zoologico Italiano. Supplemento* 8, 1 (1976).
- Zinner, D., Atickem, A., Beehner, J. C., Bekele, A., Bergman, T. J., Burke, R., Dolotovskaya, S., Fashing, P. J., Gippoliti, S., Knauf, S., Knauf, Y., Mekonnen, A., Moges, A., Nguyen, N., Stenseth, N. C., & Roos, C. (2018). Phylogeography, mitochondrial DNA diversity, and demographic history of geladas (*Theropithecus gelada*). *PloS one*, 13(8), e0202303. <https://doi.org/10.1371/journal.pone.0202303>
- Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American journal of physical anthropology*, 150(1), 133–140. <https://doi.org/10.1002/ajpa.22185>

Conclusions

Reticulate Evolution in Baboons

The results of the analyses reported in this thesis confirm previous concerns on the presence of real reproductive barriers among forms of baboons that are clearly distinct from morphological, behavioural, and genetic matters. As such, the biological concept of species proves its limitations in its adoption to these animals. On the other hand, the phylogenetic species concept seems more appropriate, as these forms are indeed distinct from a phylogenetic point of view. However, the occurrence of past and present gene flow complicates the analysis of genetic data as markers as mitochondrial DNA produces non-monophyletic species clades possibly due to phenomena such as nuclear swamping, as previously suggested (Zinner et al. 2013; Roos et al. 2021; Kopp et al. 2023). Even in terms of morphology, we have seen cases (e.g. *Gorongosa chacma*) where it appears to be present a continuum between forms rather than sharp divides, with intermediate populations that are effectively a mosaic of traits from both parental lines (Martinez et al. 2019).

The results presented here thus highlight this interconnectedness between populations, suggesting a refined interpretation of the tree of life metaphor, making that tree resemble more of an intricate bush or even a web of life (Arnold 2009). In this web, each node is connected to another, which in turn is linked to yet others. Each node represents the convergence of distinct lineages that have reconnected over time. This is what emerges at both the population and species levels.

Web of Lineages: Gene Flow and Hybridisation in Baboons

In Southern Africa, populations of *chacma* baboons appear interconnected through gene flow. Geographically closer populations are also more closely related in terms of allele sharing and have diverged more recently, while the opposite is true for populations that are geographically further apart (e.g. clinal variation). Gene flow has not only connected populations but also transferred genomic portions across neighbouring species. These flows are diluted across sources to such an extent that, in the far South of the continent, populations exhibit barely detectable signals of introgression. However, as one moves northwards, the signal of introgression increases, becoming extremely evident in populations in the Northern range (Zambia and Mozambique). In these populations, the signals of hybridisation are clear. Genetically, the *Gorongosa* population is a *chacma*-yellow mosaic, the result of the union of these two lineages in the past. In Zambia, the existence of confirmed *chacma*/*kinda* hybrids provides a clear example of ongoing hybridisations. What emerges is a complex history of interconnections and continuous flows of genetic material enhancing within species *chacma* population structure with yellow and *kinda* baboons.

In East Africa, the situation for olive baboons appears quite similar. Here too, there seems to be an interconnection between populations from Tanzania to Ethiopia. These two areas seem to be where local *P. anubis* populations experienced hybridisation. Differently from the South of Africa where the hybridisations were between *Papio* species phylogenetically related (all three belonging to the Southern clade, as noted in the introduction of this thesis) and the yellow and kinda baboons were quite similar - so much so that until relatively recently, one was considered a subspecies of the other - in Eastern Africa, hybridisation occurred with hamadryas baboons in the Northern population and with yellow baboons in the Southern population. Yellow and hamadryas baboons are well distinct from each other morphologically, genetically, and behaviorally. They even belong to the two different clades into which baboon species are divided, with hamadryas and olive baboons being part of the Northern clade and yellow baboon part of the Southern clade. This differential hybridisation among the various populations of *P. anubis* results in significant heterogeneity within the species, with populations at the Northern and Southern extremes of the range carrying genetic components associated with two distinct species, contributing to a structured population.

In Ethiopia, olive baboons have also hybridised with a third species from another genus, *Theropithecus gelada*. The population showing the greatest signal of allele sharing appears to be the Southern gelada population in the Arsi Mountains, South of the Great Rift Valley. This population seems to have diverged from the other gelada populations (particularly the central one) around 20,000 years ago and has since remained isolated from the rest of the species by a major ecological barrier, the Great Rift Valley, leading to high inbreeding and extremely low diversity. Unlike the hybridisation between baboons, the hybridisation between baboons and *Theropithecus gelada* is old; despite recent observations of such events.

Considerations on Conservation Status

The example of the Southern gelada population raises a clear question that can no longer be ignored: is it still viable to use species as conservation units, or would it be more appropriate to focus on individual populations? As is well known, as a species *Theropithecus gelada* is considered Least Concern by IUCN (Gippoliti et al. 2019). However, the population South of the Rift Valley is clearly struggling; only a few thousand individuals surviving were estimated (De Jong et al. 2023). The Arsi geladas habitat suffers from intense anthropogenic pressure and no protected areas exist in the region. The remaining mosaic forests are threatened by intensive deforestation and agricultural ploughing. Montane habitats suffer from burning during the dry season and overgrazing during the wet season (Kostin et al. 2019). For these reasons, the Arsi gelada may be considered Endangered according to IUCN criteria. In general, it seems more appropriate to evaluate each case individually, determining where intervention is necessary to preserve diversity. In cases where different populations display specific traits that make them unique and distinct from others, it would be more appropriate to treat them as conservation units separate from the species as a whole, and protected independently of their taxonomic status. As such, studies of this one are essential in clarifying the genetic profiles of populations.

Another conservation concern relates to hybrid populations. These populations are the natural product of interbreeding and are indeed something distinct from the parental forms. Where

hybridisation zones are stable, meaning where hybrid populations are stable and show no negative effects due to their nature, these populations represent something new that must be recognised as such (Fitzpatrick et al. 2015). Populations that are hybrid, such as those sampled by Chiou et al. (2021) in Zambia, or populations like those in Mozambique that show morphological and genetic signals of ancient hybridisation events, should, in my view, be considered conservation units, since they possess unique characteristics that distinguish them from the source species. However, these differences are not sufficient to classify them as an entirely new species, nor as the original species, but as something innovative. These chacma populations would indeed be part of the *P. u. griseipes* subspecies, which, as already mentioned in Chapter 1, appears to be an intermediate form between Typical chacma and yellow baboons.

In general, the principle seems valid: studying population structure and potential hybridisation to assess whether some populations possess unique traits that make them important for the species diversity, whether these traits are derived from hybridisation or other phenomena, such as isolation and different ecological conditions from the rest of the populations.

Ethical Considerations Regarding Museum Collections

The use of museum specimens collected when ethical issues in science were largely absent, requires us to take into appropriate consideration the circumstances under which such items were brought to European museums. Frequently, specimens from non-European territories were acquired in colonial times. This context was marked by local exploitation of people and their natural resources, treating them as mere properties of European powers, often trampling on the dignity of those people and their lands. I believe that scientists of the former colonising nations, and who make use of the products tainted by those practices, bear a moral obligation to acknowledge the history of how these specimens found their way to Western museums.

Museums are institutions dedicated to the preservation of memories, but they also enact narratives that evolve over time, not only according to the state of research but also as a result of changing cultural contexts, ideological perspectives, and contemporary interests. When most of the museums emerged in the second half of the XIX century, they were conceived as places for the preservation of materials for the study of specialists, as well as educational centres for the entertainment and instruction of the public (Browne 1992). The museographic narrative was initially and for a long time inspired by an evolutionist perspective and, at the same time, by the celebration of geographical and military missions carried out by various nations. In these spaces, the colonial power of the nation was thus celebrated (Browne 1992). This was also true, although more subtly, for natural history museums. As reported in Chapter 2, the natural sciences were closely linked to colonialism: although not always directly involved, they benefited from the exploitation of subjugated territories and the marginalisation of local populations (Ashby and Machin 2021; Miriti, Rawson and Mansfield 2023). Furthermore, the acquisition of zoological material was often the result of what we would today call poaching, where the term ‘collecting’ should in fact be read as ‘killing’ (Blencowe 2021; Ashby and Machin 2021). The progress of the natural sciences was a direct result of colonialism. Many naturalists joined expeditions funded by European states or institutions from colonising nations with the aim of mapping territories (as in the case of Darwin’s expedition), collecting plants, animals, and other

resources. These expeditions provided naturalists with the necessary infrastructure to explore unfamiliar areas and "discover" species that would later enrich museums, botanical gardens, and private collections in Europe. Thus, the science of natural history played an active role in the exploitation of colonial resources, from which it benefited, and still does today, often without recognising the contributions of local populations (Miriti, Rawson and Mansfield 2023).

Highlighting the past of plunder, abuse, and exploitation of the peoples and lands they inhabited is therefore important not only for the use of collections for scientific purposes but also for their contextualisation within the museums where they are housed. Ethically, it is imperative that these institutions, as some have begun to do (Ashby and Machin 2021), acknowledge the history of exploitation and looting that brought these materials into their collections, facilitating the advancement of science. This process of "decolonisation" of museum collections is necessary not only to inform the public of the stories behind these collections but also to foster critical reflection on the practices of science, which have historically excluded and marginalised local communities (Miriti, Rawson and Mansfield 2023).

The decolonisation of museum institutions does not consist of removing materials derived from colonial activity, ignoring their origin, or hiding them from the public. On the contrary, it means creating museum spaces that provide opportunities for reflection on the motivations and rhetoric used at the time in displaying objects brought to Europe to document the conquest of territories. It also involves illustrating the perspectives and rhetoric behind the scientific missions promoted in colonial territories that led to the enrichment of institutions with exotic plant and animal species, as well as other artefacts such as ethnographic and anthropological ones. Colonial collections must be exhibited in order to be discussed, deconstructed, and re-narrated. A simple removal could, in fact, lead to a new oblivion, avoiding a public confrontation with the colonial past and preventing it from becoming lasting social memory. Decolonisation means deconstructing to reconstruct a new and different museum that encourages visitors to question and reflect. Decolonial practice in museums consists of addressing historical imbalances between oppressors and the oppressed in the narratives represented within these institutions. Such activities are based on dismantling systemic hierarchies in which European narratives have traditionally been considered superior to others, for example, by showing how the contributions of colonised peoples were sidelined in order to elevate European achievements. It also includes exploring how museum collections were assembled based on power imbalances, as a product of empire (Ashby and Machin 2021). As it is necessary to tell the stories of the collections, it is equally necessary to recount how the museums themselves were involved in colonial practices in the past (for example by commissioning expeditions) and how they benefited from them. As can be clearly understood, this is not about rewriting history, as is often claimed, but simply about telling history without the rhetoric of the victor over the vanquished, from another perspective, recounting the horrors of colonisation and the suffering it caused. In this way, even visiting a natural history museum can effectively highlight that part of history which is often forgotten, perhaps out of shame or because it was not one's own people who were subjugated, nor one's own land that was plundered. Moreover, even today, the West still benefits from that cultural legacy.

Future Directions

Returning to the main topics of this thesis, several routes for future study can be identified. First and foremost, it would be essential to continue studying chacma baboons by increasing the number of populations sampled, particularly from areas that have yet to be explored, but which appear to be of great interest. For instance, it would be particularly valuable to obtain samples from the region of Luangwa Valley in Zambia, where it seems that three species may converge, or at the very least, where the geographical proximity between them is greatest. Another promising region is Western and Southern Mozambique, where chacma populations are present, to fill the gap about the sources of gene flow with chacma from this area. Lastly, sampling the Western portion of Southern Africa, particularly Southern Namibia where Jolly identified the Ruacana subspecies, and obtaining samples from Southern Angola where the range of the chacma baboon ends and that of the kinda baboon begins, would also be crucial (FIGURE 0.1).

In addition, obtaining samples of kinda baboons from Angola, Congo, and Luangwa Valley in Zambia, and samples of yellow baboons from Northern Mozambique and Zambia, would be valuable. These populations could facilitate the identification of the population source of hybridisation events described in this thesis, better characterise hybridisations, and potentially reveal additional ones (for example in Angola; Koop et al. 2023). Furthermore, it would be fascinating to delve deeper into the genomes to identify the specific introgression regions as the result of these hybridisations and assess their functional significance. For example, we have seen that the Gorongosa chacma baboon is a genetic mosaic—what does this mean from a functional perspective? Does being a combination of chacma and yellow baboons, in terms of gene expression, result in any differences compared to Cape chacma baboons, for instance?

Another significant research direction involves a dual approach linking population genetics with morphology, which will need to address the question: How does genetic variation affect morphology? This approach could prove highly valuable when applied to other species of baboons, such as in the hybridisation zone between yellow and olive baboons. Here, this line of inquiry is particularly feasible as cranial scans and the genomes of numerous individuals from the Texas SNPRC colony, which houses both yellow and olive baboons, are readily available. The hybridisation zone between olive and hamadryas baboons also warrants further investigation, particularly by adding samples from populations such as olive baboons from Eritrea and Sudan, as the museum data used in this thesis suggest some genetic exchange in this region with hamadryas baboons.

Concerning the use of museum data, it may prove useful to revisit these analyses once a phased reference genome becomes available, enabling the more accurate imputation of museum genomes, despite their low coverage. Finally, it would be extremely valuable to obtain samples of hamadryas baboons from Arabia, to complete studies based on mitochondrial fragments that have sought to explore the out-of-Africa expansion of *Papio* (Koop et al. 2014).

Regarding *Theropithecus gelada*, sampling in the Tigray region could prove highly insightful (Girmay and Tesfay 2020). As mentioned earlier, a population exists there that could revolutionise our understanding of gelada populations, traditionally divided into Northern, central, and Southern groups. Should genomic data support this, it would lead to a necessary revision

of subspecies designations, potentially recognising Western, Eastern, and Southern subspecies, or even elevating the Southern population to species status. The issue concerning the Southern subspecies could also be resolved by integrating published work with morphological analyses. To date, no studies have examined the cranial features of this group of animals and compared them with those of *T. g. obscurus* and *T. g. gelada*. Finally, it would be useful to further investigate the functional aspects related to the number of amylase gene copies, to verify whether this is reflecting differences in dietary habits.

Thus, in light of the results presented and the considerable amount of research that remains to be done, it is clear that the evolutionary history of baboons and their hybridisation is far from fully understood. However, this thesis has laid the groundwork for future studies. While much remains to be explored, a small step forward has been made. This small step could prove significant for someone else in advancing knowledge in this field, progressing with small but meaningful steps. This is the nature of scientific progress. After all, “Rome wasn’t built in a day”.

References

- Arnold, M. L. (2009). Reticulate evolution and humans: origins and ecology. *Oxford University Press*.
- Ashby, J., and Machin, R. 2021. Legacies of colonial violence in natural history collections. *Journal of Natural Science Collections*. 8. pp.44-54
- Blencowe, M. (2021). Gone: A search for what remains of the world's extinct creatures. *Leaping Hare Press*.
- Browne, J.E. (1992). A Science of Empire: British Biogeography before Darwin. *Revue d'histoire des sciences* 45, 453-475. <https://doi.org/10.3406/rhs.1992.4244>
- Chiou, K. L., Bergey, C. M., Burrell, A. S., Disotell, T. R., Rogers, J., Jolly, C. J., & Phillips-Conroy, J. E. (2021). Genome-wide ancestry and introgression in a Zambian baboon hybrid zone. *Molecular ecology*, 30(8), 1907–1920. <https://doi.org/10.1111/mec.15858>
- De Jong, Y.A., Moges, A., Jenner, T., et al. *Theropithecus gelada* ssp. *Arsi*. The IUCN Red List of Threatened Species 2023 [provisional].
- Fitzpatrick, B. M., Ryan, M. E., Johnson, J. R., Corush, J., and Carter, E. T. (2015). Hybridization and the Species Problem in Conservation. *Current Zoology* 61 (1): 206–16.
- Gippoliti, S., Mekonnen, A., Burke, R., Nguyen, N. & Fashing, P.J. (2019). *Theropithecus gelada*. The IUCN Red List of Threatened Species.

- Girmay, T., & Tesfay, H. (2020). New geographical range of geladas (*Theropithecus gelada*) in Tigray region, Northern Ethiopia. *African Journal of Ecology*, 59(1), 277–280. <https://doi.org/10.1111/aje.12818>
- Kopp, G. H., Roos, C., Butynski, T. M., Wildman, D. E., Alagaili, A. N., Groeneveld, L. F., & Zinner, D. (2014). Out of Africa, but how and when? The case of hamadryas baboons (*Papio hamadryas*). *Journal of human evolution*, 76, 154–164. <https://doi.org/10.1016/j.jhevol.2014.08.003>
- Kopp, G. H., Sithaldeen, R., Trede, F., Grathwol, F., Roos, C., & Zinner, D. (2023). A Comprehensive Overview of Baboon Phylogenetic History. *Genes*, 14(3), 614. <https://doi.org/10.3390/genes14030614>
- Kostin, D., Kasso, M., Komarova, V., Martynov, A., Gromov, A., Alexandrov, D., Bekele, A., Zewdie, C., Bryja, J. & Lavrenchenko, L. (2019). Taxonomic and genetic diversity of rodents from the Arsi Mountains (Ethiopia). *Mammalia*, 83(3), 237–247. <https://doi.org/10.1515/mammalia-2017-0135>
- Martinez, F. I., Capelli, C., Ferreira da Silva, M. J., Aldeias, V., Alemseged, Z., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Habermann, J. M., Lüdecke, T., Madiquida, H., Mathe, J., Negash, E., Paulo, L. M., Pinto, M., Stalmans, M., Tátá, F., & Carvalho, S. (2019). A missing piece of the *Papio* puzzle: Gorongosa baboon phenostucture and intrageneric relationships. *Journal of human evolution*, 130, 1–20. <https://doi.org/10.1016/j.jhevol.2019.01.007>
- Miriti, M. N., Rawson, A. J., & Mansfield, B. (2023). The history of natural history and race: Decolonizing human dimensions of ecology. *Ecological applications : a publication of the Ecological Society of America*, 33(1), e2748. <https://doi.org/10.1002/eap.2748>
- Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in *Papio* baboons and their phylogeographic implications. *American journal of physical anthropology*, 174(3), 407–417. <https://doi.org/10.1002/ajpa.24186>
- Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American journal of physical anthropology*, 150(1), 133–140. <https://doi.org/10.1002/ajpa.22185>

Supplementary Table 1

Information on samples used in Chapter 1						
Newly generated samples						
Sample	Genus	Species	Country	Locality	Population	Genome coverage
MT18	<i>Papio</i>	<i>ursinus</i>	SouthAfrica	Cape	Cape	16.95
NE712562	<i>Papio</i>	<i>ursinus</i>	Namibia	Tsaobis National Park	Namibia	37.83
NE736543	<i>Papio</i>	<i>ursinus</i>	Namibia	Tsaobis National Park	Namibia	35.11
PR4	<i>Papio</i>	<i>ursinus</i>	SouthAfrica	Cape	Cape	18.59
MSNG32161	<i>Papio</i>	<i>ursinus</i>	SouthAfrica	-	Transvaal_museum	0.055
MA2	<i>Papio</i>	<i>ursinus</i>	SouthAfrica	Limpopo	Limpopo	40.49
MA3	<i>Papio</i>	<i>ursinus</i>	SouthAfrica	Limpopo	Limpopo	24.46
BF59	<i>Papio</i>	<i>ursinus</i>	Mozambique	Gorongosa Natational Park	Mozambique	0.006
BF221	<i>Papio</i>	<i>ursinus</i>	Mozambique	Catapù Reserve Forest	Mozambique	0.012
BF270	<i>Papio</i>	<i>ursinus</i>	Mozambique	Catapù Reserve Forest	Mozambique	0.057
BF312	<i>Papio</i>	<i>ursinus</i>	Mozambique	Gorongosa Natational Park	Mozambique	0.011
BF315	<i>Papio</i>	<i>ursinus</i>	Mozambique	Gorongosa Natational Park	Mozambique	0.035
Samples from NCBI						
Sample	Genus	Species	Country	Locality	Population	Project
-	<i>Mandrillus</i>	<i>leucophaeus</i>	-	-	Drill	PRJNA785018
HAP	<i>Papio</i>	<i>cynocephalus</i>	Kenya	Amboseli Nat. Park	PcynK	Wall et al. 2019
1X2304	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X0102	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1_3548	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X0110	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X1734	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X2049	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X2208	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
1X3576	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	PcynK	Robinson et al. 2019
PD_0214	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Ruaha	PcynTE	Sørensen et al. 2023
PD_0215	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Ruaha	PcynTE	Sørensen et al. 2023
PD_0226	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Udzungwa	PcynTE	Sørensen et al. 2023
PD_0227	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Udzungwa	PcynTE	Sørensen et al. 2023
PD_0231	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Mahale	PcynTW	Sørensen et al. 2023
PD_0232	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Mahale	PcynTW	Sørensen et al. 2023
PD_0233	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Katavi	PcynTW	Sørensen et al. 2023
PD_0234	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Katavi	PcynTW	Sørensen et al. 2023
PD_0235	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Selous	PcynTE	Sørensen et al. 2023
PD_0237	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Selous	PcynTE	Sørensen et al. 2023
PD_0662	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Issa Valley	PcynTW	Sørensen et al. 2023
PD_0746	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Mikumi National Park	PcynTE	Sørensen et al. 2023
PD_0770	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Mikumi National Park	PcynTE	Sørensen et al. 2023
SRR7717284	<i>Papio</i>	<i>hybrid</i>	Zambia	Top Musa	Hybrid	Chiou et al. 2021
SRR7717285	<i>Papio</i>	<i>hybrid</i>	Zambia	Top Musa	Hybrid	Chiou et al. 2021
SRR7717286	<i>Papio</i>	<i>hybrid</i>	Zambia	Top Musa	Hybrid	Chiou et al. 2021
SRR7717287	<i>Papio</i>	<i>hybrid</i>	Zambia	Top Musa	Hybrid	Chiou et al. 2021
SRR7717288	<i>Papio</i>	<i>hybrid</i>	Zambia	Malala Camp	Hybrid	Chiou et al. 2021
SRR7717289	<i>Papio</i>	<i>hybrid</i>	Zambia	North Nkala Road	Hybrid	Chiou et al. 2021
SRR7717290	<i>Papio</i>	<i>hybrid</i>	Zambia	Musa Bridge	Hybrid	Chiou et al. 2021
SRR7717291	<i>Papio</i>	<i>hybrid</i>	Zambia	Musa Bridge	Hybrid	Chiou et al. 2021
SRR7717292	<i>Papio</i>	<i>hybrid</i>	Zambia	Musa Bridge	Hybrid	Chiou et al. 2021
SRR7717311	<i>Papio</i>	<i>hybrid</i>	Zambia	Chunga	Hybrid	Chiou et al. 2021
SRR7717323	<i>Papio</i>	<i>hybrid</i>	Zambia	North Nkala Road	Hybrid	Chiou et al. 2021
SRR7717324	<i>Papio</i>	<i>hybrid</i>	Zambia	North Nkala Road	Hybrid	Chiou et al. 2021
SRR7717327	<i>Papio</i>	<i>hybrid</i>	Zambia	Nanzhila Plains	Hybrid	Chiou et al. 2021
SRR7717330	<i>Papio</i>	<i>hybrid</i>	Zambia	North Nkala Road	Hybrid	Chiou et al. 2021
SRR7717331	<i>Papio</i>	<i>hybrid</i>	Zambia	North Nkala Road	Hybrid	Chiou et al. 2021
SRR7717332	<i>Papio</i>	<i>hybrid</i>	Zambia	NanzhilaPlains	Hybrid	Chiou et al. 2021
SRR7717333	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021

SRR7717335	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717336	<i>Papio</i>	<i>hybrid</i>	Zambia	Malala Camp	Hybrid	Chiou et al. 2021
SRR7717339	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717342	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717343	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717345	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717363	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717364	<i>Papio</i>	<i>hybrid</i>	Zambia	Chunga	Hybrid	Chiou et al. 2021
SRR7717372	<i>Papio</i>	<i>hybrid</i>	Zambia	Top Musa	Hybrid	Chiou et al. 2021
SRR7717373	<i>Papio</i>	<i>hybrid</i>	Zambia	Musa Bridge	Hybrid	Chiou et al. 2021
SRR7717374	<i>Papio</i>	<i>hybrid</i>	Zambia	Ngoma Airstrip	Hybrid	Chiou et al. 2021
SRR7717375	<i>Papio</i>	<i>hybrid</i>	Zambia	Malala Camp	Hybrid	Chiou et al. 2021
SRR7717377	<i>Papio</i>	<i>hybrid</i>	Zambia	Mwengwa Rapids	Hybrid	Chiou et al. 2021
SRR7717378	<i>Papio</i>	<i>hybrid</i>	Zambia	Mwengwa Rapids	Hybrid	Chiou et al. 2021
SRR7717379	<i>Papio</i>	<i>hybrid</i>	Zambia	Mwengwa Rapids	Hybrid	Chiou et al. 2021
SRR7717380	<i>Papio</i>	<i>hybrid</i>	Zambia	Mwengwa Rapids	Hybrid	Chiou et al. 2021
SRR7717382	<i>Papio</i>	<i>hybrid</i>	Zambia	Mwengwa Rapids	Hybrid	Chiou et al. 2021
PD_0749	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Sørensen et al. 2023
PD_0751	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Sørensen et al. 2023
PD_0758	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Sørensen et al. 2023
34449	<i>Papio</i>	<i>kindae</i>	Zambia	Kafue National Park	PkinK	Rogers et al. 2019
34472	<i>Papio</i>	<i>kindae</i>	Zambia	Kafue National Park	PkinK	Rogers et al. 2019
34474	<i>Papio</i>	<i>kindae</i>	Zambia	Kafue National Park	PkinK	Rogers et al. 2019
SRR7717274	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717275	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717276	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717277	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717278	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717279	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717280	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717281	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717282	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717283	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717293	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717294	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717295	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717296	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717297	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717298	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717299	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717300	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717301	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717302	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717303	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717304	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717305	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717306	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717307	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717308	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717309	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717310	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717312	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717313	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717314	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717315	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717316	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717317	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717318	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717319	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717320	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021

SRR7717321	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717322	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717334	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717340	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717361	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717362	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717365	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717366	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717367	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717368	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717369	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717370	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717371	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717383	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717384	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717385	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717386	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717387	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717388	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717389	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717390	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717391	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717392	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717393	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717394	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717396	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717397	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717398	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717399	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717400	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717401	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
SRR7717402	<i>Papio</i>	<i>kindae</i>	Zambia	Chunga	PkinC	Chiou et al. 2021
ERP119178	<i>Papio</i>	<i>ursinus</i>	Lesotho	Ha Makotoko	Ancient_Lesotho	Mathieson et al. 2020
BF186	<i>Papio</i>	<i>ursinus</i>	Mozambique	Gorongosa Natational Park	Mozambique	Santander et al. 2022
BB1	<i>Papio</i>	<i>ursinus</i>	Mozambique	Gorongosa Natational Park	Mozambique	Santander et al. 2022
PD_0694	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia	Sørensen et al. 2023
PD_0695	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia	Sørensen et al. 2023
PD_0693	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia	Sørensen et al. 2023
PD_0692	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia	Sørensen et al. 2023
SRR7717325	<i>Papio</i>	<i>ursinus</i>	Zambia	Nanzhila Plains	Zambia_RAD	Chiou et al. 2021
SRR7717326	<i>Papio</i>	<i>ursinus</i>	Zambia	Nanzhila Plains	Zambia_RAD	Chiou et al. 2021
SRR7717328	<i>Papio</i>	<i>ursinus</i>	Zambia	Nanzhila	Zambia_RAD	Chiou et al. 2021
SRR7717329	<i>Papio</i>	<i>ursinus</i>	Zambia	North Nkala Road	Zambia_RAD	Chiou et al. 2021
SRR7717338	<i>Papio</i>	<i>ursinus</i>	Zambia	Ngoma Airstrip	Zambia_RAD	Chiou et al. 2021
SRR7717341	<i>Papio</i>	<i>ursinus</i>	Zambia	Ngoma Airstrip	Zambia_RAD	Chiou et al. 2021
SRR7717344	<i>Papio</i>	<i>ursinus</i>	Zambia	Ngoma Airstrip	Zambia_RAD	Chiou et al. 2021
SRR7717346	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia_RAD	Chiou et al. 2021
SRR7717347	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia_RAD	Chiou et al. 2021
SRR7717348	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia_RAD	Chiou et al. 2021
SRR7717349	<i>Papio</i>	<i>ursinus</i>	Zambia	Dendro Park	Zambia_RAD	Chiou et al. 2021
SRR7717350	<i>Papio</i>	<i>ursinus</i>	Zambia	Lower Zambezi National Park	Zambia_RAD	Chiou et al. 2021
SRR7717351	<i>Papio</i>	<i>ursinus</i>	Zambia	Lower Zambezi National Park	Zambia_RAD	Chiou et al. 2021
SRR7717353	<i>Papio</i>	<i>ursinus</i>	Zambia	Lower Zambezi National Park	Zambia_RAD	Chiou et al. 2021
SRR7717354	<i>Papio</i>	<i>ursinus</i>	Zambia	Lower Zambezi National Park	Zambia_RAD	Chiou et al. 2021
SRR7717355	<i>Papio</i>	<i>ursinus</i>	Zambia	Lower Zambezi National Park	Zambia_RAD	Chiou et al. 2021
SRR7717356	<i>Papio</i>	<i>ursinus</i>	Zambia	Choma	Zambia_RAD	Chiou et al. 2021
SRR7717358	<i>Papio</i>	<i>ursinus</i>	Zambia	Choma	Zambia_RAD	Chiou et al. 2021
SRR7717359	<i>Papio</i>	<i>ursinus</i>	Zambia	Choma	Zambia_RAD	Chiou et al. 2021
SRR7717360	<i>Papio</i>	<i>ursinus</i>	Zambia	Ngoma Airstrip	Zambia_RAD	Chiou et al. 2021

Supplementary Table 2

Information on samples used in Chapter 2							
Newly generated samples							
Sample	Genus	Species	Country	Locality	Cluster	Date of collection	Genome coverage
MSG4085	<i>Papio</i>	<i>anubis</i>	Ethiopia	Scio-a	Et	1870-1880	0.058
MSG4086	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Scio-a Fekerie-Ghem	hamadryas	-	0.136
MSG4088	<i>Papio</i>	<i>anubis</i>	Ethiopia	Scio-a	Et	1870-1880	0.11
MSG4089	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Scio-a Fekerie-Ghem	hamadryas	-	0.098
MSG4092	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Scio-a	hamadryas	-	0.126
MSG4093	<i>Papio</i>	<i>cynocephalus</i>	Somalia	-	hamadryas	1891-1993	0.095
MSG4094	<i>Papio</i>	<i>hamadryas</i>	-	-	hamadryas	1888	0.099
MSG31312	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Gaarre, Dancalia	hamadryas	1928	0.124
MSG34764	<i>Papio</i>	<i>doguera</i>	Ethiopia	Asile	Et	1939	0.05
MSG34765	<i>Papio</i>	<i>doguera</i>	Ethiopia	Monti Amar	Et	1939	0.088
MSG36417	<i>Papio</i>	<i>hamadryas</i>	Somalia	-	hamadryas	1956	0.099
MSG42504	<i>Papio</i>	<i>anubis</i>	DRC	Alto uelle	Ce	1927	0.119
MSNM75	<i>Papio</i>	<i>hamadryas</i>	Somalia	Boga Aléd	hamadryas	1953	0.06
MSNM4280	<i>Papio</i>	<i>hamadryas</i>	Eritrea	-	Et	-	0.053
Samples from NCBI							
Sample	Genus	Species	Country	Locality	Cluster	Origin	Project
30877	<i>Papio</i>	<i>anubis</i>	Kenya	Arberdale	Ce	Modern	Rogers et al. 2019
30977	<i>Papio</i>	<i>anubis</i>	Kenya	Arberdale	Ce	Modern	Rogers et al. 2019
90456	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90458	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90460	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90461	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90462	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90463	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
90464	<i>Papio</i>	<i>anubis</i>	Kenya	Masai Mara	Ce	Modern	Wall et al. 2016
1_1979	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Ce	Modern	Robinson et al. 2019
1x0812	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Et	Modern	Robinson et al. 2019
1x0832	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Ce	Modern	Robinson et al. 2019
1x1700	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Ce	Modern	Robinson et al. 2019
1x1765	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Ce	Modern	Robinson et al. 2019
1x1939	<i>Papio</i>	<i>anubis</i>	Kenya	SNPRC (founder)	Ce	Modern	Robinson et al. 2019
1x4384	<i>Papio</i>	<i>cynocephalus</i>	Kenya	SNPRC (founder)	yellow	Modern	Robinson et al. 2019
PD_0200	<i>Papio</i>	<i>anubis</i>	Tanzania	Serengeti	Ce	Modern	Sørensen et al. 2023
PD_0201	<i>Papio</i>	<i>anubis</i>	Tanzania	Serengeti	Ce	Modern	Sørensen et al. 2023
PD_0202	<i>Papio</i>	<i>anubis</i>	Tanzania	Serengeti	Ce	Modern	Sørensen et al. 2023
PD_0203	<i>Papio</i>	<i>anubis</i>	Tanzania	Serengeti	Ce	Modern	Sørensen et al. 2023
PD_0204	<i>Papio</i>	<i>anubis</i>	Tanzania	Arusha	Ar	Modern	Sørensen et al. 2023
PD_0205	<i>Papio</i>	<i>anubis</i>	Tanzania	Arusha	Ar	Modern	Sørensen et al. 2023
PD_0206	<i>Papio</i>	<i>anubis</i>	Tanzania	Arusha	Ar	Modern	Sørensen et al. 2023
PD_0207	<i>Papio</i>	<i>anubis</i>	Tanzania	Arusha	Ar	Modern	Sørensen et al. 2023
PD_0208	<i>Papio</i>	<i>anubis</i>	Tanzania	Ngorongoro	So	Modern	Sørensen et al. 2023
PD_0209	<i>Papio</i>	<i>anubis</i>	Tanzania	Ngorongoro	So	Modern	Sørensen et al. 2023
PD_0210	<i>Papio</i>	<i>anubis</i>	Tanzania	Ngorongoro	So	Modern	Sørensen et al. 2023
PD_0211	<i>Papio</i>	<i>anubis</i>	Tanzania	Ngorongoro	So	Modern	Sørensen et al. 2023
PD_0219	<i>Papio</i>	<i>anubis</i>	Tanzania	Gombe	Ce	Modern	Sørensen et al. 2023
PD_0220	<i>Papio</i>	<i>anubis</i>	Tanzania	Gombe	Ce	Modern	Sørensen et al. 2023
PD_0221	<i>Papio</i>	<i>anubis</i>	Tanzania	Gombe	Ce	Modern	Sørensen et al. 2023
PD_0222	<i>Papio</i>	<i>anubis</i>	Tanzania	Gombe	Ce	Modern	Sørensen et al. 2023
PD_0239	<i>Papio</i>	<i>anubis</i>	Tanzania	Lake Manyara	So	Modern	Sørensen et al. 2023
PD_0240	<i>Papio</i>	<i>anubis</i>	Tanzania	Lake Manyara	So	Modern	Sørensen et al. 2023
PD_0241	<i>Papio</i>	<i>anubis</i>	Tanzania	Lake Manyara	So	Modern	Sørensen et al. 2023
PD_0242	<i>Papio</i>	<i>anubis</i>	Tanzania	Lake Manyara	So	Modern	Sørensen et al. 2023
PD_0244	<i>Papio</i>	<i>anubis</i>	Tanzania	Tarangire	Ta	Modern	Sørensen et al. 2023
PD_0265	<i>Papio</i>	<i>anubis</i>	Tanzania	Tarangire	Ta	Modern	Sørensen et al. 2023

PD_0266	<i>Papio</i>	<i>anubis</i>	Tanzania	Tarangire	Ta	Modern	Sørensen et al. 2023
PD_0267	<i>Papio</i>	<i>anubis</i>	Tanzania	Tarangire	Ta	Modern	Sørensen et al. 2023
PD_0720	<i>Papio</i>	<i>anubis</i>	Ethiopia	Gog	Et	Modern	Sørensen et al. 2023
PD_0722	<i>Papio</i>	<i>anubis</i>	Ethiopia	Gog	Et	Modern	Sørensen et al. 2023
PD_0725	<i>Papio</i>	<i>anubis</i>	Ethiopia	Gog	Et	Modern	Sørensen et al. 2023
PD_0735	<i>Papio</i>	<i>anubis</i>	Ethiopia	Gog	Et	Modern	Sørensen et al. 2023
SAMEA112757472	<i>Papio</i>	<i>anubis</i>	Sudan	Shiulluk Islands	Ce	Museum	Grathwol et al. 2023
SAMEA112757473	<i>Papio</i>	<i>anubis</i>	Sudan	Sennaar	Et	Museum	Grathwol et al. 2023
SAMEA112757477	<i>Papio</i>	<i>anubis</i>	South Sudan	Nagishot	Ce	Museum	Grathwol et al. 2023
aDNA99	<i>Papio</i>	<i>anubis</i>	DRC	Kivu	Ce	Museum	Roos et al. 2020
aDNA98	<i>Papio</i>	<i>anubis</i>	DRC	Kivu	Ce	Museum	Roos et al. 2020
aDNA97	<i>Papio</i>	<i>anubis</i>	Niger	Aïr	Ce	Museum	Roos et al. 2020
aDNA96	<i>Papio</i>	<i>anubis</i>	CAR	Bamingui	Ce	Museum	Roos et al. 2020
aDNA95	<i>Papio</i>	<i>anubis</i>	Niger	Aïr	Ce	Museum	Roos et al. 2020
aDNA94	<i>Papio</i>	<i>anubis</i>	Sierra Leone	Simeria Chiefdom	Ce	Museum	Roos et al. 2020
PD_0770	<i>Papio</i>	<i>cynocephalus</i>	Tanzania	Mikumi National Park	yellow	Modern	Sørensen et al. 2023
aDNA640	<i>Papio</i>	<i>doguera</i>	Chad	Tibesti	Ce	Museum	Roos et al. 2020
PD_0706	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Sørensen et al. 2023
PD_0713	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Sørensen et al. 2023
PD_0716	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Sørensen et al. 2023
PD_0718	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Sørensen et al. 2023
SAMN20949839	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Chiou et al. 2022
SAMN20949842	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Chiou et al. 2022
SAMN20949843	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Chiou et al. 2022
SAMN20949844	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	hamadryas	Modern	Chiou et al. 2022
SAMEA112757486	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Harar	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757474	<i>Papio</i>	<i>hamadryas</i>	Eritrea	Keren, Bogos	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757478	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Djeddah-Valley	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757475	<i>Papio</i>	<i>hamadryas</i>	Sudan	Sawakin	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757476	<i>Papio</i>	<i>hamadryas</i>	-	-	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757480	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Harar	Et	Museum	Grathwol et al. 2023
SAMEA112757482	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Adua	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757481	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Dire Dawa	hamadryas	Museum	Grathwol et al. 2023
SAMEA112757479	<i>Papio</i>	<i>hamadryas</i>	Djibuti	Djibouti-harbour	hamadryas	Museum	Grathwol et al. 2023
aDNA100	<i>Papio</i>	<i>hamadryas</i>	Somalia	[Upper Sheikh]	hamadryas	Museum	Roos et al. 2020
-	<i>Mandrillus</i>	<i>leucophaeus</i>	-	-	Drill	Modern	PRJNA785018

Supplementary Table 3

Information on samples used in Chapter 3						
Newly generated samples						
Sample	Genus	Species	Country	Locality	Population	Genome coverage
A10	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Lega Aba Boge-Lega Adem	Southern	12
A28	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Gado-Goro	Southern	30
TGLZ1	<i>Theropithecus</i>	<i>gelada</i>	-	Wilhelma Zoo Stuttgart	Zoo	33.65
LACZ1	<i>Lophocebus</i>	<i>aterrimus</i>	-	Gaia Zoo Kerkrade	Lophocebus	34.98
Samples from NCBI						
Sample	Genus	Species	Country	Locality	Population	Project
SAMN20949847	<i>Theropithecus</i>	<i>gelada</i>	-	Lincoln Children's Zoo	Zoo	Chiou et al. 2022
SAMN20949848	<i>Theropithecus</i>	<i>gelada</i>	-	Lincoln Children's Zoo	Zoo	Chiou et al. 2022
SAMN20949849	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949850	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949851	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949852	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949853	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949854	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949855	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949856	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Zoo	Zoo	Chiou et al. 2022
SAMN20949857	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Wild Animal Park	Zoo	Chiou et al. 2022
SAMN20949858	<i>Theropithecus</i>	<i>gelada</i>	-	San Diego Wild Animal Park	Zoo	Chiou et al. 2022
SAMN20949859	<i>Theropithecus</i>	<i>gelada</i>	-	Bronx Zoo	Zoo	Chiou et al. 2022
SAMN20949860	<i>Theropithecus</i>	<i>gelada</i>	-	Bronx Zoo	Zoo	Chiou et al. 2022
SAMN20949861	<i>Theropithecus</i>	<i>gelada</i>	-	Bronx Zoo	Zoo	Chiou et al. 2022
SAMN20949862	<i>Theropithecus</i>	<i>gelada</i>	-	Virginia Zoological Park	Zoo	Chiou et al. 2022
SAMN20949863	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Guassa	Central	Chiou et al. 2022
SAMN20949864	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Guassa	Central	Chiou et al. 2022
SAMN20949865	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Guassa	Central	Chiou et al. 2022
SAMN20949866	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Chenek	Northern	Chiou et al. 2022
SAMN20949867	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Chenek	Northern	Chiou et al. 2022
SAMN20949868	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Chenek	Northern	Chiou et al. 2022
SAMN20949869	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Chenek	Northern	Chiou et al. 2022
SAMN20949870	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Chenek	Northern	Chiou et al. 2022
SAMN20949871	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949872	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949873	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949874	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949875	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949876	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949877	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949878	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949879	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949880	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949881	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949882	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949883	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949884	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949885	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949886	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949887	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949888	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949889	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949890	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN09197221	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949891	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949892	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949893	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949894	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949895	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949896	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022

SAMN20949897	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949898	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949899	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949900	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949901	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949902	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949903	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949904	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949905	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949906	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949907	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949908	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949909	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949910	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949911	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949912	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949913	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949914	<i>Theropithecus</i>	<i>gelada</i>	Ethiopia	Sankaber	Northern	Chiou et al. 2022
SAMN20949842	<i>Papio</i>	<i>hamadryas</i>	Ethiopia	Filoha	Baboon	Chiou et al. 2022
PD_0720	<i>Papio</i>	<i>anubis</i>	Ethiopia	Gog	Baboon	Sørensen et al. 2023
-	<i>Mandrillus</i>	<i>leucophaeus</i>	-	-	Drill	PRJNA785018

Supplementary Table 4

Monthly precipitation for five locations (Indetu, Wollo, Guassa, Sankaber and Gich) over the years 2006-2017 recovered from CRU-TS 4.06 (downscaled with WorldClim 2.1)												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Guassa_2006	4,20	5,00	102,00	58,60	46,60	47,50	414,20	318,50	154,00	49,60	2,60	21,00
Sankaber_2006	2,10	1,00	12,50	16,10	80,50	59,40	299,20	250,60	103,50	40,80	6,40	6,60
Gich plateau_2006	1,50	0,50	10,60	13,30	137,40	70,00	337,90	299,90	107,50	36,10	9,00	4,50
Indetu_2006	12,60	36,40	148,50	142,30	98,90	63,20	149,70	147,30	133,10	114,30	24,30	33,40
Koshemene_2006	11,60	14,90	119,60	62,60	59,40	54,10	212,50	137,40	101,10	44,60	9,70	50,00
Guassa_2007	18,20	18,80	43,70	50,00	44,20	112,00	383,70	303,10	132,40	26,60	3,40	0,50
Sankaber_2007	7,20	2,50	9,50	26,50	34,20	110,90	329,70	239,00	76,40	14,80	4,90	3,10
Gich plateau_2007	4,20	1,00	9,10	24,60	48,00	130,10	416,40	288,40	72,10	13,20	6,10	2,80
Indetu_2007	36,00	40,60	77,10	120,50	126,60	96,30	131,90	148,80	130,40	88,00	11,10	3,70
Koshemene_2007	58,10	30,80	65,10	68,40	54,20	87,90	186,80	142,20	79,20	19,10	6,60	4,10
Guassa_2008	5,70	4,10	9,80	25,30	71,50	54,70	342,50	325,90	116,90	26,80	60,70	7,20
Sankaber_2008	5,70	1,00	11,00	38,20	66,90	102,50	265,30	233,80	79,60	17,50	23,90	7,80
Gich plateau_2008	3,70	0,50	10,30	48,40	113,10	103,00	281,40	283,70	90,00	12,10	24,20	5,70
Indetu_2008	16,50	22,30	29,00	91,00	146,70	57,00	122,00	169,30	121,50	87,30	154,00	20,50
Koshemene_2008	22,70	11,10	14,40	53,70	59,30	68,20	182,00	143,90	80,80	25,10	84,80	38,80
Guassa_2009	20,00	18,00	23,70	28,40	23,20	81,50	290,40	359,20	68,20	52,20	11,10	9,00
Sankaber_2009	4,10	15,60	4,70	13,00	9,20	96,30	260,30	233,90	57,10	34,30	4,10	11,30
Gich plateau_2009	2,50	7,30	4,90	13,00	15,00	94,40	317,20	242,50	57,00	25,80	4,50	7,60
Indetu_2009	53,70	24,30	56,80	101,80	97,40	67,80	111,20	177,00	89,60	116,40	35,70	18,30
Koshemene_2009	51,10	72,10	35,60	33,90	25,30	70,00	152,90	169,00	51,30	42,20	12,20	44,80
Guassa_2010	4,80	69,50	88,60	57,50	54,80	64,80	308,10	297,10	71,30	30,90	21,00	12,30
Sankaber_2010	10,60	5,90	11,70	30,00	56,50	83,40	259,40	217,60	100,50	25,70	10,70	6,80
Gich plateau_2010	7,00	2,50	12,30	37,30	92,90	83,10	314,80	249,90	114,30	22,10	14,30	4,50
Indetu_2010	14,20	119,80	115,80	145,30	106,80	64,60	117,00	143,70	72,60	85,50	66,90	28,70
Koshemene_2010	36,50	129,70	93,60	69,40	56,00	65,60	162,00	132,50	48,70	27,00	33,20	56,30
Guassa_2011	11,50	6,80	28,60	107,40	54,80	81,90	308,10	361,20	121,40	17,60	19,50	4,50
Sankaber_2011	5,40	1,00	11,10	16,90	64,70	81,20	221,00	178,90	88,20	23,60	14,00	1,00
Gich plateau_2011	3,40	0,50	10,30	13,30	99,90	82,70	249,60	193,60	89,60	23,00	20,30	0,70
Indetu_2011	26,90	21,20	47,40	252,20	107,10	79,10	117,00	162,60	130,00	54,10	48,00	14,30
Koshemene_2011	36,10	14,50	38,40	128,50	56,00	82,70	162,00	145,80	87,50	20,10	26,30	12,50
Guassa_2012	11,50	29,60	63,00	73,20	54,80	65,60	329,00	297,10	114,50	30,90	10,50	5,00
Sankaber_2012	5,40	5,70	11,00	16,70	65,90	82,50	276,90	216,50	68,40	22,80	21,60	8,80
Gich plateau_2012	3,40	2,40	10,30	13,20	99,30	87,10	301,90	248,50	58,20	18,90	28,20	5,90
Indetu_2012	26,90	55,80	88,50	179,30	107,00	65,20	117,00	143,70	117,90	85,20	35,10	15,60
Koshemene_2012	36,10	56,30	82,50	88,10	55,90	66,30	187,90	132,50	85,50	27,00	27,70	32,50
Guassa_2013	19,90	1,40	64,00	73,70	125,60	58,60	310,80	308,40	71,60	89,40	13,70	0,50
Sankaber_2013	4,40	1,00	3,20	74,60	35,30	74,40	184,90	284,40	56,30	62,70	6,00	1,00
Gich plateau_2013	2,50	0,50	3,30	75,70	47,20	72,90	210,30	347,60	59,90	54,40	9,70	0,70
Indetu_2013	33,20	20,10	93,70	140,90	165,40	67,30	109,10	151,80	101,30	161,40	47,50	3,70
Koshemene_2013	44,70	7,60	54,00	142,20	69,80	73,20	143,40	141,60	51,90	61,00	9,60	4,10
Guassa_2014	3,30	17,20	59,00	49,20	65,90	63,10	336,70	310,40	161,80	69,00	5,50	2,60
Sankaber_2014	5,80	10,70	41,20	41,10	83,70	45,20	266,00	283,90	133,20	48,90	6,70	10,20
Gich plateau_2014	3,90	4,90	44,70	46,00	121,70	49,60	305,30	334,50	153,00	41,10	8,30	7,30
Indetu_2014	11,60	22,40	74,30	147,30	117,40	65,10	122,80	139,30	135,90	158,00	17,90	4,10
Koshemene_2014	21,20	52,60	122,50	74,00	78,90	40,80	164,40	144,80	105,60	48,40	11,20	23,00
Guassa_2015	1,30	2,60	43,60	6,00	83,20	94,50	195,40	347,20	108,90	12,40	11,00	5,30
Sankaber_2015	1,50	1,00	12,00	9,90	117,30	58,30	121,50	247,00	93,90	18,50	13,60	17,30
Gich plateau_2015	1,10	0,50	11,40	14,50	206,30	60,30	152,00	285,20	101,80	18,60	16,70	12,50
Indetu_2015	19,80	51,40	77,10	50,50	140,80	83,60	115,30	157,60	97,40	89,50	30,00	14,10
Koshemene_2015	9,80	16,20	76,30	10,60	80,90	72,40	106,20	147,20	72,80	17,00	22,60	40,70
Guassa_2016	21,80	22,80	34,40	100,00	134,60	82,30	315,30	366,90	111,00	12,80	6,90	1,10
Sankaber_2016	6,20	1,50	11,60	29,90	82,10	105,40	304,20	251,80	100,90	25,20	4,40	1,10
Gich plateau_2016	3,60	0,70	12,60	26,70	109,90	97,30	363,60	280,80	119,00	31,50	5,30	0,80
Indetu_2016	61,50	59,70	63,60	234,10	302,20	58,90	110,70	170,60	99,60	71,70	19,70	7,60
Koshemene_2016	73,60	28,30	58,90	110,00	133,00	78,50	165,30	156,50	75,10	17,20	10,00	7,40
Guassa_2017	0,60	26,10	68,60	52,60	76,40	67,70	298,40	386,20	184,20	53,70	1,70	0,50
Sankaber_2017	1,30	6,70	13,70	30,60	104,10	105,90	239,00	288,70	83,40	39,60	1,70	1,00
Gich plateau_2017	1,10	2,80	13,70	29,60	146,80	109,90	271,60	313,80	81,50	30,90	2,40	0,70
Indetu_2017	6,60	47,50	78,10	96,40	118,40	63,60	119,90	148,90	145,20	119,80	15,40	8,70
Koshemene_2017	4,50	55,10	74,70	63,70	68,40	70,70	160,80	170,90	90,60	49,00	4,00	4,70

Appendix 1

Gelada genomes highlight events of gene flow, hybridisation and local adaptation that track past climatic changes

Matteo Caldon¹  | Giacomo Mutti^{1,2,3} | Alessandro Mondanaro⁴ | Hiroo Imai⁵  | Takayoshi Shotake⁶ | Gonzalo Oteo Garcia^{1,7,8}  | Gurja Belay⁹ | Jordi Morata¹⁰ | Jean-Rémi Trotta¹⁰ | Francesco Montinaro^{11,12} | Spartaco Gippoliti^{13,14} | Cristian Capelli^{1,15} 

¹Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

²Barcelona Supercomputing Centre (BSC-CNS), Barcelona, Spain

³Institute for Research in Biomedicine (IRB Barcelona), the Barcelona Institute of Science and Technology, Barcelona, Spain

⁴Department of Earth Sciences, University of Florence, Florence, Italy

⁵Center for the Evolutionary Origins of Human Behavior, Kyoto University, Inuyama, Aichi, Japan

⁶Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

⁷Centre for Palaeogenetics, Stockholm, Sweden

⁸Department of Archaeology and Classical Studies, Stockholm University, Stockholm, Sweden

⁹Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia

¹⁰Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain

¹¹Department of Biology-Genetics, University of Bari, Bari, Italy

¹²Institute of Genomics, University of Tartu, Tartu, Estonia

¹³IUCN/SSC Primate Specialist Group, Rome, Italy

¹⁴Società Italiana per la Storia Della Fauna "G. Altobello", Rome, Italy

¹⁵Department of Biology, University of Oxford, Oxford, UK

Correspondence

Cristian Capelli, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11/a, 43124 Parma, Italy.

Email: cristian.capelli@unipr.it

Funding information

Leakey Foundation; Italian Ministry for Education, University and Research; 'La Caixa' Foundation, Grant/Award Number: LCF/BQ/DI22/11940014; Fondazione con il Sud, Grant/Award Number: 2018-PDR-01136; the European Regional Development Fund, Grant/Award Number: 2014-2020.4.01.16-0030; Italian Ministry of University and Research, Grant/Award Number: 2022P2ZER

Handling Editor: Alana Alexander

Abstract

Theropithecus gelada, the last surviving species of this genus, occupy a unique and highly specialised ecological niche in the Ethiopian highlands. A subdivision into three geographically defined populations (Northern, Central and Southern) has been tentatively proposed for this species on the basis of genetic analyses, but genomic data have been investigated only for two of these groups (Northern and Central). Here we combined newly generated whole genome sequences of individuals sampled from the population living south of the East Africa Great Rift Valley with available data from the other two gelada populations to reconstruct the evolutionary history of the species. Integrating genomic and paleoclimatic data we found that gene-flow across populations and with *Papio* species tracked past climate changes. The isolation and climatic conditions experienced by Southern geladas during the Holocene shaped local diversity and generated diet-related genomic signatures.

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

© 2024 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

KEYWORDS

evolutionary history, gene-flow, paleoclimate, population genomics, *Theropithecus gelada*

1 | INTRODUCTION

Theropithecus gelada is the only extant species belonging to the primate genus, *Theropithecus* (Delson, 1993). This species occupies a unique ecological niche, the Afroalpine grasslands of the Ethiopian Highlands, and is specialised for altitudes ranging from 2000 to 4500 m (Belay & Shotake, 1998). Geladas are the sole graminivorous primates, consuming a grass-based diet similarly to ungulates (Iwamoto, 1993a). Their diet comprises primarily graminoid leaves (up to 90% of the ingested food; Dunbar, 1977). The second most common dietary component is underground food (UF: tubers, roots and corms), a nutritional fall-back resource rich in starch and mostly consumed when the grass is scarce as the result of decreased rainfall (Iwamoto, 1993a, 1993b; Jarvey et al., 2018).

On the basis of morphological traits, two major subspecies have been described: *T. g. gelada* (Rüppell, 1835) and *T. g. obscurus* Heuglin, 1863. The two subspecies are usually associated respectively with the northern region and the central region of the Northern Ethiopian plateau (de Beaux, 1925; Gippoliti, 2010; Girmay & Tesfay, 2020). However, the absence of clear-cut differences, particularly when considering the variation in coat pigmentation, and the apparent absence of a major barrier between the two taxa has led to scepticism about the taxonomic validity of the subspecies classification (Yalden et al., 1976, 1996). Mitochondrial DNA (mtDNA) phylogenetic analyses are broadly congruent with the Northern-Central subdivision first proposed by de Beaux (de Beaux, 1925; Shotake et al., 2016; Zinner et al., 2018). The geographical range of modern-day geladas was extended in 1990 when a population was found south of the Great Rift Valley in the South-eastern plateau, in the Arsi mountains (Mori & Belay, 1990). In 2016 the subspecies *Theropithecus gelada arsi* Shotake Saijuntha Agatsuma et Kawamoto, 2016, was tentatively proposed for this clearly isolated population on the basis of coat coloration and molecular data (Belay & Shotake, 1998; Gippoliti, 2010; Mori & Belay, 1990; Shotake et al., 2016).

Given the debate about the geographical distribution of the three gelada subspecies and concerns about their taxonomic validity, in this work, we will name the three gelada demes as Northern, Central and Southern geladas, as done by others in previous investigations (Chiou et al., 2022; Trede et al., 2020; Zinner et al., 2018) acknowledging that some degree of overlap is present between these geographically described units and the *T. g. gelada*, *T. g. obscurus* and *T. g. arsi* taxa (cf. Figure 1; Gippoliti, 2010; Shotake et al., 2016).

Early molecular investigations showed significant differentiation between populations across the Ethiopian Rift Valley (Central and Southern geladas), with the two groups being characterised by similar levels of within-group variation (Belay & Shotake, 1998).

Notably, the mitochondrial lineages of the Southern geladas of the Arsi mountains are identified either as basal to all the other geladas (Central and Northern; Shotake et al., 2016) or are placed within the Northern variation, sister to a Northern subclade (Zinner et al., 2018). Despite these differences in tree topologies, the investigations conducted by Zinner et al. (2018) and Shotake et al. (2016) are concordant in supporting a relatively recent Time to the Most Recent Common Ancestor (TMRCA) of gelada mtDNA (670–400 thousand years ago, kya), and in showing Southern geladas as characterised by the lowest mtDNA diversity across the three gelada populations. Interestingly, no lineages are shared between geladas living in the three regions (with the exception of a likely mislabelled sample, as suggested by Zinner et al. (2018)), suggesting deep population structure among the three groups, at least for the maternal lineages. Microsatellite data from a subset of the samples analysed by Zinner et al. (2018) confirmed a

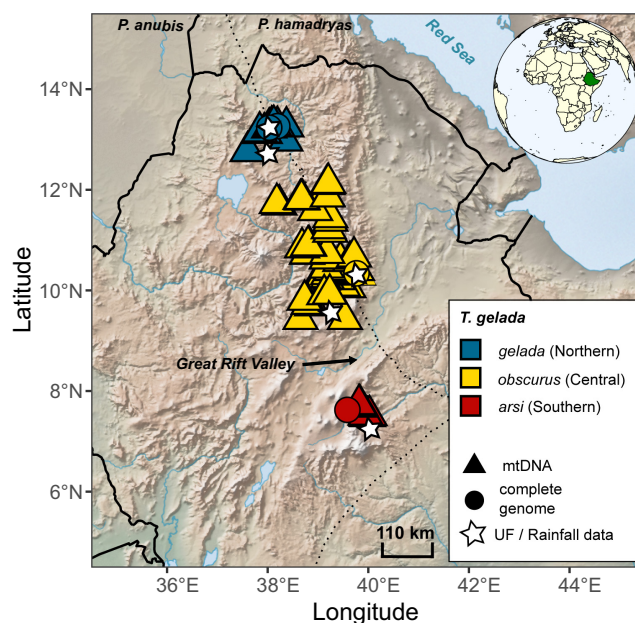


FIGURE 1 Geographical origin in the Ethiopian highlands of *Theropithecus gelada* samples analysed in this work. Colours refer to source populations (Northern, Central and Southern) and shapes identify the data available for the sample (Complete genome sequence and mitochondrial DNA only). The dotted line provides an approximate indication of the distribution of *P. hamadryas* (right) and *P. anubis* (left) in the area of interest, as indicated in Sørensen et al. (2023). The stars provide the location of provenience of the five datasets with UF consumption and rainfall data (from top to bottom: Northern: Sankaber (Jarvey et al., 2018) and Gich (Woldegeorgis & Bekele, 2015), Central: Guassa (Fashing et al., 2014) and Wollo (Kifle & Bekele, 2021), and Southern: Indetu (Abu et al., 2018)). Prepared using Natural Earth. Free vector and raster map data @[naturalearthdata.com](https://www.naturalearthdata.com).

three-deme structure for geladas (Trede et al., 2020), the genetic distance, measured using F_{ST} , between Southern geladas and the other two populations being two-three times larger than the distance between Central and Northern (Trede et al., 2020). More recently, whole genome sequence data of individuals from two regions in the Northern (Sankaber and Chenek), one in the Central (Guassa) and five US zoos were released (Chiou et al., 2022). The published preliminary population genomic analysis of this dataset confirmed the subdivision between Northern and Central geladas, zoo specimens (ultimate origin of these animals or their ancestors was unknown) showing mostly a Central gelada ancestry component. The Northern-Central separation also emerged at karyotypic level as a different number of chromosomes were suggested for individuals sampled in the two regions (42 and 44 respectively). Centric fission of chromosome 7 was detected in Northern gelada, resulting in two acrocentric chromosomes, 7a and 7b, and $2n=44$ karyotype. Such a macroscopic genomic difference was interpreted as evidence for putative incipient speciation as hybrids between lineages with heterozygous karyotypes might have reduced reproductive success (King, 1995). Genome-wide data were also used to estimate how the effective population size of Northern and Central geladas changed in the past. Both showed an almost steady decrease over the last million years; notably, Northern gelada and Central gelada diverged in their pattern 500 kya, suggesting this as the initial temporal point of separation of the two groups (Chiou et al., 2022).

The Ethiopian plateau is part of the East African Rift System, an area profoundly affected by environmental instability during the late Quaternary (Foerster et al., 2012). Geochemical, physical and biological indicators testify that Ethiopia's climate regime was controlled by the intensification of the dry-wet-dry cycles during the last glacial period, potentially linked to the change in orbitally forced radiation (Trauth et al., 2003). Species response to such changes usually includes habitat tracking towards different latitudes, altitudes or the population concentrated within refugia (Keppel et al., 2012; Raia et al., 2012). The gelada is probably no exception, since the species is currently confined to the Ethiopian Highland at high altitudes whereby the three recognised populations occupy different areas of the plateau (Keppel et al., 2012; Raia et al., 2012).

Assessing the ecological similarities and differences among gelada populations is therefore fundamental to understanding the effect of climate on their current distribution. Despite its relevance, climate variability has been explored only in relation to the ecological diversification and fate of different members of the genus *Theropithecus* (Dunbar, 1993; Pigkford, 1993), while recent work has investigated how future climatic changes might alter the spatial distribution of geladas (Ahmed et al., 2023). Nevertheless, an extensive approach to derive a climatic-driven spatial distribution of gelada in the past is yet to be applied. To this aim, we used Species Distribution Models (SDMs), a powerful tool widely used in ecology, biogeography, biodiversity conservation and natural resources management (Guillera-Arroita et al., 2015). SDMs are

further indicated to complement fossil and genetic evidence by providing spatial prediction of past species distribution (Svenning et al., 2011).

In addition, hybridisation, another process potentially influenced by climatic variation, has been suggested between geladas and members of the genus *Papio*, possibly having occurred more than once, on the basis of mobile elements being shared across the genomes of baboon species and geladas (Walker et al., 2019). However, besides the rejection of a phylogenetic tree that excluded gene flow between geladas and *Papio* (Santander et al., 2022), no formal testing for hybridisation event(s) has been performed so far. Mating events between geladas and baboons and the possible presence of hybrids have been reported both in captivity and in the wild (Dunbar & Dunbar, 1974; Jolly et al., 1997; Markarjan et al., 1974).

Overall, the emerging consensus based on DNA supports at least three evolutionary units for *T. gelada*, with a robust parallelism between genetics and geography, possibly the result of their adaptation to the montane grassland environment (Chiou et al., 2022). However, the impact of climatic changes on geladas current geographical distribution and past population dynamics has been poorly explored, particularly considering the intense climatic oscillations of the Late Pleistocene. Here we integrate population genomic data with explicit modelling of species habitat suitability to better clarify the linkage between species ecological preferences and demographic dynamics, with the aim of generating an integrated reconstruction of the evolutionary history of *Theropithecus gelada*.

As is often the case in biological conservation, reaching a consensus on taxonomy and conservation units by integrating as many different data sets as possible (Gippoliti, 2020; Groves et al., 2017) may be crucial for immediate management actions for *Theropithecus*. This may include the identification of priority sites for research and conservation in Ethiopia, the recognition of the significance of ex-situ population programs and the future planning for re-stocking and/or reintroduction actions. We note that while the conservation status of the overall species is *Least Concern* (LC; Gippoliti et al., 2019) the geographical features of the Ethiopian plateau can endanger the survival of populations due to isolation and small census size, as well exemplified by Southern geladas.

2 | MATERIALS AND METHODS

2.1 | Sampling and samples dataset

Gelada blood samples from different locations in Ethiopia were collected by two of the authors (G.B. and T.S.) during 1978–1997 (more details can be found in Belay & Shotake, 1998; Shotake et al., 2016). The guidelines for field research of the Kyoto University Primate Research Institute and the rules and regulations of the Ethiopian government at the time were followed during blood collection (Belay & Shotake, 1998). From this collection, DNA samples from the Arsi (Southern) population were evaluated for their compatibility with whole-genome sequencing, and

two samples, one from Gado-Goro (A28) and one from Lega Aba Boge-Lega Adem (A10), were selected to be whole-genome sequenced. Sample A28 was sequenced to a depth of 30x, while sample A10 reached a median coverage of 12x, the sequencing being completed by EuroFin Japan. In addition, we sequenced a gelada sample provided via the European Association of Zoos and Aquaria (EAZA) by the Wilhelma Zoo Stuttgart and processed by the EAZA biobank hub at the Leibniz Institute for Zoo and Wildlife Research (HJZ14-00788), here referred to as TGLZ1 (the ultimate geographical origin of all captive animals included in this study and their ancestors is unknown). The sample TGLZ1 was sequenced at the Centro Nacional de Análisis Genómico in Barcelona, reaching a final coverage of 33.65x. We also generated whole genome sequence data from a sample of *Lophocebus aterrimus* provided by the Gaia Zoo Kerkrade and processed by the EAZA hub at Copenhagen Zoo. This sample was sequenced at the Centro Nacional de Análisis Genómico in Barcelona with a mean coverage of 34.98x. The taxonomic position of the genus *Lophocebus*, sister to *Papio*, makes it possible to highlight *Papio*-specific evolutionary dynamics.

In this study, we also included all the samples of *T. gelada* and one of *Papio hamadryas* (hamadryas baboon) from Chiou et al. (2022), one sample of *Papio anubis* (olive baboon) from Sørensen et al. (2023) and one sample of *Mandrillus leucophaeus* (drill) recovered from PRJNA785018 BioProject in NCBI (Table S1). We focused our attention on the two *Papio* species present in Ethiopia (*P. anubis* and *P. hamadryas*) and known to have been in contact with geladas (Dunbar & Dunbar, 1974; Jolly et al., 1997). In this study, we used the term *hybridisation* to refer to interbreeding between *T. gelada* and *Papio*, at any point in time in their evolutionary history, and admixture for gene flow between gelada populations.

2.2 | Genome mapping

Reads were mapped with *BWA-MEM2* v.2.2.1 (Vasimuddin et al., 2019), to the reference assembly genome *Tgel_1.0* (Chiou et al., 2022). Since there was no Y chromosome in this reference genome, we added it from the *Papio* reference assembly genome *Panubis1* (Batra et al., 2020). Reads from (Chiou et al., 2022) were trimmed with *fastp* v.0.23.2 (Chen et al., 2018) in order to remove Nextera adapters. We then marked the duplicate reads with *Picard MarkDuplicates* version 2.8.1 (<http://broadinstitute.github.io/picard/>) and filtered the result using the options -q 10, -F 1292 and -f 2 of *samtools view* (Li et al., 2009).

To avoid some analyses being influenced by the reference genome (in particular, those involving comparisons with other genera, see later; Günther & Nettelblad, 2019), we additionally mapped *Papio*, *Lophocebus*, *Mandrillus* and a subset of *Theropithecus* samples (11 high-coverage gelada individuals belonging to the four localities of sampling—for the details see Table S1) to the *Macaca mulatta* reference genome (*Mmul_10*; Warren et al., 2020), following the same procedure described above.

2.3 | Mitochondrial DNA analysis

We downloaded the publicly available mitogenomes from Chiou et al. (2022) (GenBank id from [MZ907630](#) to [MZ907677](#)), and the fragments generated in previous analyses of geladas mitochondrial DNA variation (Shotake et al., 2016; Zinner et al., 2018; [LC018113](#) to [LC018133](#) and [MH634017](#) to [MH634077](#) respectively). As outgroups, we selected male samples from close genera with available genomic reads in order to be able to include them also in the Y chromosome phylogeny, see Table S1. The mitogenomes of the newly sequenced samples, including outgroups with no available mitogenomes, were assembled with *GetOrganelle* v1.7.5 (Jin et al., 2020), with parameters -R 10 -s NC_020006.2.fa -F animal_mt --fast; then the resulting scaffolds were filtered by length (if it was between 16,000 and 17,000 nucleotides) and manually rotated. All the complete mitogenomes were aligned with *MAFFT* v7.490 (Katoh & Standley, 2013) with the --auto option. Further, the *cyt-b* fragments were added with the --addfragments option of *MAFFT* to the mitogenomes alignment of *T. gelada* and *Papio* samples from Zinner et al., 2013 and Roos et al., 2021 (*Papio* samples were not included in the tree here presented). The D-loop was manually removed from the alignment. Both Maximum Likelihood mitochondrial trees were computed with *IQ-Tree* v2.2.0 (Minh et al., 2020). The best mutational model was inferred with *ModelFinder* (Kalyaanamoorthy et al., 2017) and branch support values were estimated with 1000 Ultrafast bootstraps. Bayesian inference and divergence time dating were done with *BEAST* v.2.7.1 (Bouckaert et al., 2019). TPM2u+F+I+G4, as implemented in *SSM* v.1.2.0 *BEAST2* package (Bouckaert & Xie, 2017), was selected as the substitution model based on the ML *ModelFinder* results. The trees were modelled under a Yule process with a strict clock with a uniform distribution. The monophyly of the 2 *Macaca* samples was enforced to root the tree with two fossil calibrated nodes: *Papio-Theropithecus* (5 MYA ± 0.5 SD, normal distribution, as in Dolotovskaya et al. (2017) and the root node (5.3 MYA ± 0.5 SD, lognormal distribution as suggested in Pozzi and Penna 2022). Two replicates were performed for 50MLn generations, sampling every 1000 trees.

2.4 | Y chromosome analysis and Y STR genotyping

We called the variants on the Y chromosome with *bcftools* v.1.15.1 (Li, 2011) by setting the ploidy to 1. Then we filtered constant sites, indels and variants in repetitive regions with *VCFTools* v.0.1.16 (Danecek et al., 2011). Further, after checking the resulting VCF, variants with minimum read depth (DP) lower than 250, maximum DP greater than 600, minimum mapping quality less than 55 and absent in less than 10 samples were removed. Finally, the resulting variants were thinned out by keeping only variants at least 100bp apart. We then converted the final VCF in an alignment with *vcf-2phylip* v.2.0 ([10.5281/zenodo.2540861](#)). Then we used *BEAST2* to obtain a dated phylogeny with a TPM2 substitution model with four gamma categories, a strict clock. We used as a calibration point the

speciation between gelada and *Papio* as Normal $5MLn \pm 0.5SD$. We ran four MCMC iterations of 10MLn samples each.

We applied TYPESTer to the genomes of the gelada samples from the Arsi (Southern) region and Leibniz Institute here analysed to genotype the set of Y-STRs previously recovered from other gelada samples (Mutti et al., 2023). The STRs were assembled in haplotypes and merged with the haplotypic data previously generated (Table S3). Haplotypes were used to reconstruct a phylogenetic tree using the Bruvo (Bruvo et al., 2004) distance from poppr R package (Kamvar et al., 2014) followed by UPGMA clustering, as described in Mutti and colleagues (Mutti et al., 2023).

2.5 | Variant calling and the final dataset

We called the autosomal chromosome variants, using the GATK version 4.2.4.1 (McKenna et al., 2010). For each sample, we generated the gVCF with *HaplotypeCaller*, and then we made the joint calls via *GenotypeGVCFs*. The VCF thus obtained was filtered with *VariantFiltration* using hard filters for SNPs ('QD <2.0 || MQ <40.0 || FS >60.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0'). The final VCF was therefore composed of 70 individuals of *T. gelada* and 9,244,659 variants. Using *PLINK* v.1.9 (Chang et al., 2015), all sites from the sex chromosomes and with a minor allele frequency (MAF) below 0.05 were removed, keeping only biallelic SNPs filtered using the *--snps-only* option.

2.6 | Whole-genome analysis

2.6.1 | Heterozygosity, RoH and kinship

To calculate the proportion of heterozygous sites (HET) we used *PLINK --het*. This command estimated the observed O(HOM) homozygous genotype counts for each sample and the total number of variable sites (N_Sites), from which we derived HET per individual with the formula:

$$HET[N_SitesOHOM] / N_Sites$$

We estimated the number, the total length and the average length of runs of homozygosity (RoH) with *PLINK --homozyg* and the option *--homozyg-snp 25 --homozyg-kb 200 --homozyg-window-snp 10*. The degree of kinship among samples was tested up to the third degree using *KING* v.2.3.1 with *--kinship --degree 3* flags on the *PLINK* bed file (Manichaikul et al., 2010).

2.6.2 | PCA

We performed a preliminary survey of *T. gelada* population structure by running a principal component analysis (PCA) using the *smart-pca* function implemented in *EIGENSOFT* software 8.0.0 (Patterson et al., 2006) using default settings.

2.6.3 | D-statistics

We performed the *D-statistics* (ABBA-BABA test) using *qpDstat* (Patterson et al., 2012), with default parameters, to test for the occurrence of any imbalance in allele sharing between different populations and species in the form D(H4,H3;H2,H1) including drill (*M. leucophaeus*) as an outgroup (H4). We firstly tested whether the two Northern populations of Sankaber and Chenek (as H1 and H2) were symmetrical in their sharing with other gelada groups (Central or Southern as H3) and what was the degree of allelic sharing between the three gelada taxa (Northern, Central, Southern alternatively placed as H1, H2 or H3). Then we tested the allelic sharing of these gelada populations with *Papio* species and *Lophocebus* (Dunbar & Dunbar, 1974; Jolly et al., 1997) in the form D(Drill, Papio; Gelada2, Gelada1), D(Drill, Lophocebus; Gelada2, Gelada1), D(Drill, Gelada; Papio2, Papio1) and D(Drill, Gelada; Lophocebus, Papio), where 'Papio' were alternatively *P. anubis* or *P. hamadryas* genomes, 'Lophocebus' was the genome of *L. aterrimus* and 'Gelada' was a genome from one of the three gelada populations.

The test considers the ancestral alleles (A) and the derived alleles (B) in the genomes of the four taxa considered. Without gene flow, the two conflicting allelic patterns (ABBA and BABA) are equally frequent, and the values of D are not significantly different from zero. In the presence of gene flow an excess of one of the two allelic patterns occurs, which produces a deviation of the D from zero: positive D (excess of BABA) indicates the introgression between H1 and H3, and negative D (excess of ABBA) points to the gene flow between H2 and H3. The estimates were considered significant if $|Z|$ -score >3.

The *D-statistics* analysis involving the genera *Papio* and *Lophocebus* was conducted using reads mapped on the *Mmul_10* reference genome, to avoid a reference bias.

2.6.4 | Admixture graphs

We explored the relationships and the gene-flow events between the three gelada taxa and the species of the genus *Papio* (*P. anubis* and *P. hamadryas*) and *L. aterrimus* using *find_graphs* in R package *admixturetools2* (Maier et al., 2023). The R command *find_graphs* was run with parameters *stop_gen=10,000*, *stop_gen2=100*, *plusminus_generations=20*, *eject_f4z=2*, *diag=1e-04*, *num_start=1000*, *lsqmode=FALSE*, *resid=F*, *num_admix* between 1 and 5 and *M. leucophaeus* as an outgroup; the same command was run 10 times per number of admixture events in order to obtain the graph with the score as close as possible to the global optimum. We kept, for each edge, the three trees with the best score; we reported in the main text the tree with the greatest number of admixture events and no contributions equal to 0%. The admixture graphs analysis was conducted using reads mapped on the macaque reference genome *Mmul_10*, to avoid reference bias.

2.6.5 | PSMC

Demographic changes (effective population size history) over time were explored using the pairwise sequentially Markovian coalescent (PSMC) model (Li & Durbin, 2011). As recommended, we generated diploid sequences using *bcftools* (Li, 2011); sites with coverage greater than double the average genome-wide coverage or less than one-third were excluded. We also excluded the repetitive regions in the genomes, indicated on the NCBI website (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_003255815.1). We run PSMC with default settings on autosomal chromosomes, using a mutation rate of 0.5×10^{-8} per base pair per generation and a generation time of 11.67 years as in Chiou et al. (2022). Finally, we performed bootstrapping for each individual with 10 replicates following the instructions described on the GitHub webpage manual (<https://github.com/lh3/psmc>).

In a similar manner, this software can also be used to explore the last time of contact between populations by creating artificial F1 hybrids (*hPSMC*; Cahill et al., 2016). Following the procedure described on the GitHub webpage manual (<https://github.com/jacahill/hPSMC>), we generated from the *bam* a haploid sequence in *fasta* format of the X chromosome for each male individual. We then combined the *fasta* sequences from two different individuals into a single *.psmcfa* file, and then we applied PSMC with default settings. We use a mutation rate of 0.419×10^{-8} mutations per base pair per generation. It became necessary to recalculate the autosomal mutation rate, as the X chromosome, due to its presence in different numbers in male and female individuals, results in a different mutation rate. Therefore, the parameter was obtained by converting the autosomal mutation rate into X mutation rate as described by Li and Durbin (2011), using the male-to-female mutation ratio of *Rhesus macaque* (2.9; Wilson Sayres et al., 2011). Finally, we performed bootstrapping for each hybrid as described above.

To test if the estimates obtained with *hPSMC* are influenced by the selected reference, we performed the procedure described above using reads mapped to the two reference genomes (*Tgel_1.0* and *Mmul_10*).

2.7 | Copy number variants

We further investigated differences across gelada populations by focusing on gene content variation. Scans of gelada genomes have previously identified gene families significantly expanded in comparison to *Papio* and highlighted associated biological processes, including cell proliferation, hypoxia response and immune function (Chiou et al., 2022). Variation in the number of copies present across individuals has been shown to significantly contribute to intraspecific phenotypic diversity (Almarri et al., 2020; Montinaro & Capelli, 2020; Sudmant et al., 2015).

The gelada genomic reads mapped to *Tgel_1.0* were processed via Control-FREEC to identify variations in copy number of genomic regions across the three populations of geladas (Boeva et al., 2012).

The number of copies estimated for each locus was reported for all the individuals, clustered according to the population of provenance (Northern, Central and Southern). For each expanded region the number of individuals bearing more than two copies was counted to estimate their overall occurrence and the frequency within each population. Annotated genes overlapping these regions were also reported. KEGG enrichment analyses were conducted via the clusterProfiler v.4.0.5 R package (Wu et al., 2021).

The genomic occurrence of amylase genes in the genome of geladas was verified by searching in the reference gelada genome for matches to the sequences of the human salivary amylase alpha 1a (*AMY1A*) and both pancreatic amylases (2A and 2B). Two proteins were found, both annotated in the gelada genome as pancreatic (XP_025215097.1 and XP_025215076.1). The two genomic regions hosting these genes were identified on chromosome 1: one between positions 131,757,727–131,766,740 (LOC112607926) and the other spanning positions 131,799,832–131,808,869 (LOC112607935). Amylase gene copy number was further verified in geladas using the average coverage across the two gelada amylase genomic regions, normalising it by the average coverage across chromosome 1, in each of the gelada genomes.

We reasoned that in the presence of positive selection operating on a given haplotype, the areas surrounding the locus under selection should be characterised by a decrease in genetic diversity, the size of the chromosomal region affected by this pattern being shaped by the time since the selective pressure has been in place. In order to test the hypothesis of positive selection for the duplication of the amylase locus LOC112607926 in Southern gelada, we estimated the degree of heterozygosity present in the regions surrounding the duplication in all gelada genomes.

We partitioned the 10-Mb region centred on the *AMY* genes in windows of 10kb; starting from the VCF, for each window, we calculated the number of heterozygous sites per individual using VCFtools --het. We then calculated the average for each window for each population and normalised the results by the mean of the variable sites across the entire chromosome. Starting from the windows containing the loci, we estimated the cumulative sum of the heterozygosity in each window. In order to take into account the differences across populations in the overall degree of heterozygosity, the cumulative sums were normalised by the average of heterozygosity estimated in all the 10-kb windows of chromosome 1. Finally, we calculated the average for each population, including zoo individuals in the Central population.

2.8 | Variation in underground food consumption and rainfall

Amylase expression has been linked to variation in copy number of amylase genes as the result of differences in dietary starch prevalence across human populations, dogs and pigs (Axelsson et al., 2013; Groenen et al., 2012; Pajic et al., 2019; Perry et al., 2007). Seasonal rainfall variation has been reported to affect grass availability and

Underground Food (UF; food rich in starch whose digestion is facilitated by amylases) consumption by geladas in Sankaber, grass availability significantly related to the cumulative rainfall in the previous 1–3 months (Jarvey et al., 2018). To further investigate the UF dietary content across gelada populations, we extracted from previous publications dietary data of geladas in different locations (Figure 1). Only sites with at least 1 year of continuous data collection were included, which resulted in four datasets collected over different periods: 18 months in Kosheme, Wollo (Central; data collected from May 2015 to May 2016, plus July, September, Nov 2016 and January 2017; Kifle & Bekele, 2021), 15 months in Guassa (Central; February 2007 to April 2008; Fashing et al., 2014), 12 months in Sankaber (Northern; February 2015 to January 2016; Jarvey et al., 2018) and 12 months in Gich (Northern; May 2013 to April 2014; Woldegeorgis & Bekele, 2015). We additionally included data from the Southern gelada population of Indetu, the only site with dietary data available south of the Great Rift Valley (Abu et al., 2018). Indetu comprised a set of continuous observations over 7 months extending from August 2010 to February 2011, followed by a gap in March 2011, and two additional data points for April and May 2011. In order to provide a comparison between the five datasets we initially estimated the monthly dietary content of UF (here measured as feeding time spent collecting and processing, including chewing and swallowing—dedicated to any underground part of grass and herbs, including any unclassified underground food (Jarvey et al., 2018)) over the 6-month period shared by all five datasets (August–January). We then repeated the same estimate using a sliding window of 7 months, the longest continuous set of observations in Indetu, in the other four datasets using only data collected in subsequent months with no gaps (number of windows per population: seven for Wollo, eight for Guassa, five for Sankaber and Gich). Given the link between rainfall, grass availability and UF consumption (Jarvey et al., 2018), we collected data on monthly total rainfall for the period spanning the years sampled for dietary consumption (2006–2017) for the five locations (as measured within a 20 squared km) from CRU-TS 4.06 (Harris et al., 2020) downscaled with WorldClim 2.1 (Fick & Hijmans, 2017; Table S5). Spearman correlation coefficient r_s and associated p -value between UF consumption and rainfall for the full dataset were calculated (total number of included months across locations: 66). We repeated this analysis by alternatively considering rainfall over 1, 2 or 3 months earlier than the month taken into consideration for UF consumption (Jarvey et al., 2018), as well as including in the analysis the rainfall in the month of interest for UF consumption. Evidence of significant variation in the amount of UF consumption and rainfall across the five locations was tested using one-way ANOVA.

2.9 | Climatic niche and paleoclimate reconstructions

To estimate the climatic niches of the three gelada populations we downloaded a set of 19 current bioclimatic variables from CHELSA

database version 2.1 (Karger et al., 2017). As past environmental predictors, we used the monthly bioclimatic variables generated through the 2Ma CESM1.2 simulation (Timmermann et al., 2022) downscaled at $0.5^\circ \times 0.5^\circ$ grid resolution. The native set of predictors was subsequently converted using the 'dismo' R package (Hijmans et al., 2023) to obtain the same list of the CHELSA variables. The past bioclimatic dataset covers the last 2Ma at 1 kya temporal resolution. Since paleoclimate simulations in Timmermann et al. (2022) did not model the diurnal cycle, we excluded the annual mean diurnal range (BIO2) and isothermality (BIO3) from the CHELSA original set of variables, reducing the number of bioclimatic variables to 17 (Table S6). Lastly, both current and past variables were projected on the Mollweide coordinate reference system. For SDMs calibration, CHELSA variables were rasterised at 1-km spatial resolution, while past climatic predictors were obtained at 50-km spatial resolution.

To quantify ecological similarity among the extant gelada populations, we calculated the niche overlap between them using their modern occurrences, due to the lack of *T. gelada* fossil records (Delson, 1993; Getahun et al., 2023).

Consequently, we used the bioclimatic values derived from CHELSA variables and the gelada occurrences compiled by Ahmed et al. (2023). We further filtered this data by removing duplicate records for cell removing multiple occurrence points within $1\text{ km} \times 1\text{ km}$ grid area according to the CHELSA variables spatial resolution. Niche overlap was calculated by adopting the approach described by Broennimann et al. (2011). Accordingly, PCA was used to decompose the environmental space defined by modern gelada occurrences. Occurrence data and climate conditions associated with them were projected into the PCA space, and then their densities were computed across the first two PCs relying on a kernel density estimator. Densities of occurrence and background environments were then divided by the maximum number of occurrences in all cells of the environmental space and by the number of sites with the most common climate, respectively (Broennimann et al., 2011). The process generated a density grid in the environmental space that was used to quantify niche overlap between modern niches in terms of Schoener's D index (Schoener, 1970). This latter index ranges from 0 to 1 with values of 0 indicating no niche overlap and values of 1 indicating a complete overlap. Niche overlap analyses were applied using the 'ecospat' R package (Di Cola et al., 2017). To define the environmental space for gelada, the spatial polygon which enclosed all the occurrences of the *Theropithecus* lineage (including the extinct *T. darthi* and *T. oswaldi*) was chosen as background environment since we have no a priori idea about the 'accessible area' (the geographical area that was potentially accessible given the species ecological tolerances and dispersal abilities) within *T. gelada* dispersal range. Then, we restricted this area to the elevation range where geladas currently are known to occur. Specifically, we downloaded current global topography from the CHELSA database in order to exclude cells with altitude values outside the 95% of the total elevation range associated with current *T. gelada* data points following the

approach described by Ahmed et al. (2023). Lastly, we stacked the 17 bioclimatic variables, and we cropped them by using the background environment extent. Eventually, we grouped all climatic values to verify for potential multicollinearity among variables. The full set of 17 bioclimatic variables was sub-selected considering a pairwise correlation coefficient of $r > |.75|$ (Zuur et al., 2010) through the 'usdm' R package (Naimi et al., 2014). After this procedure, we retained seven predictors: Mean Temperature of Wettest Quarter (BIO8), Mean Temperature of Driest Quarter (BIO9), Mean Temperature of Coldest Quarter (BIO11), Precipitation of Driest Month (BIO14), Precipitation Seasonality (Coefficient of Variation) (BIO15), Precipitation of Warmest Quarter (BIO18) and Precipitation of Coldest Quarter (BIO19).

Occurrence records and climatic values were projected into a PCA space to decompose the entire environmental space, then their densities were computed across the first two PCs relying on a kernel density estimator. Densities of occurrence and background environments were divided by the maximum number of occurrences in any cell of the environmental space and by the number of sites with the most common climate, respectively (Broennimann et al., 2011). The density grid in the environmental space was used to compute the niche overlap. We chose the Schoener's D index as a metric for niche comparison (Schoener, 1970). Schoener's D metric can measure the degree of niche overlap between different species or subspecies, from low ($D=0$) to perfect ($D=1$), taking into account both the climatic conditions where the species were present and the background environments.

Since there is no fossil record for the *T. gelada* and because of the very low sample size of the single living population, we chose to calibrate a SDM by using the climatic values associated with all modern occurrences and to 10,000 data points randomly picked from the background environment. For the SDM, we adopted an ensemble forecasting approach applying the functionalities described in the 'biomod2' R package (Thuiller et al., 2009). Specifically, models were trained by testing four different algorithms: Generalised Linear Model (GLM), Generalised Boosting Model (GBM), Random Forest (RF) and Maximum Entropy (MaxEnt). The predictive accuracy of models was assessed adopting a bootstrap cross-validation scheme. Data were split into 70%–30% training/testing samples and the entire procedure was repeated 10 times. To evaluate the model accuracy, we calculated the area under the operating characteristic curve (AUC; Swets, 1988). Model averaging was performed by weighting the individual model projections by their AUC values and averaging the result (Marmion et al., 2008) after avoiding poorly calibrated models with an AUC < 0.7.

Lastly, we obtained SDMs projections at each 1000 years (ky) for the last 130 kya (that is from the last interglacial to recent), focusing on temporal windows highlighted as of interest in relation to the timing of events of gene-flow and hybridisation and avoiding to consider temporal periods too different from the Late Quaternary in terms of climatic regimes.

3 | RESULTS

3.1 | Uniparental markers

We recovered mitochondrial genomes from published and newly generated whole genome sequences and reconstructed the phylogenetic relationships among the assembled mitogenomes (Figure 2a). The topology places the Southern geladas as deeply divergent, sister to the clade including both Northern and Central geladas. Central geladas are monophyletic but nested within a cluster of Northern gelada mitogenomes, sister to an additional Northern branch. The two northern branches mirror the N1 and N2 groups reported by Zinner et al. (2018), while the C2 branch, but not the C1, is possibly unsampled by currently available mitogenomes (Figure 2a; Figure S1a). The mitogenome topology is more similar to the one reported by Shotake et al. (2016) than the one presented by Zinner et al. (2018), the absence in Shotake et al. (2016) of one of the two Northern mtDNA branches (N2) possibly due to unsampled lineages (Figure S1a). The two Southern gelada mitogenomes are identical and include the previously reported haplotype h01/21, the most frequent haplotype in the Southern gelada population (Shotake et al., 2016; Zinner et al., 2018). The dating of the node grouping all geladas at 0.49 Mya (0.38–0.60) overlaps previous estimates based on shorter mtDNA regions (Shotake et al., 2016; Zinner et al., 2018; Table S2).

Y chromosome SNPs were used to generate a phylogenetic tree. Similarly, to the mtDNA results, Y chromosome sequences clustered according to the population of provenance (Northern, Central and Southern), with Southern geladas forming a clade separated from the other two. However, differently from the mtDNA data, Central and Northern geladas formed two separate sister clades (Figure 2b). Samples from zoos group in a single clade with Central gelada lineages. No further structure was detected when the two different sampling locations of the Northern gelada population were taken into consideration. The overall TMRCA for the gelada Y chromosome phylogeny was dated to 0.24 Mya (0.18–0.30 Mya; Table S2). We additionally genotyped previously described Y-STR markers in the two Southern geladas and the Stuttgart Zoo sample here presented and, using previously genotyped gelada samples, generated a phylogenetic tree (Mutti et al., 2023) (Table S3; Figure S1b). The main features of the recovered topology were in agreement with the results based on whole Y chromosome sequence data: Y-STR haplotypes clustered in accordance to population of provenance and Southern haplotypes separated from a cluster comprising Central and Northern haplotypes. The two Southern Y haplotypes differed for both SNPs and STRs.

We noted that the TMRCA of the mitochondrial DNA phylogeny was about twice the age of those recovered from the Y chromosome (Figure 2a,b). Substantial differences in TMRCA between the Y chromosome and mtDNA have been reported for gorillas and orangutans, but not for chimpanzees and humans (Hallast et al., 2016).

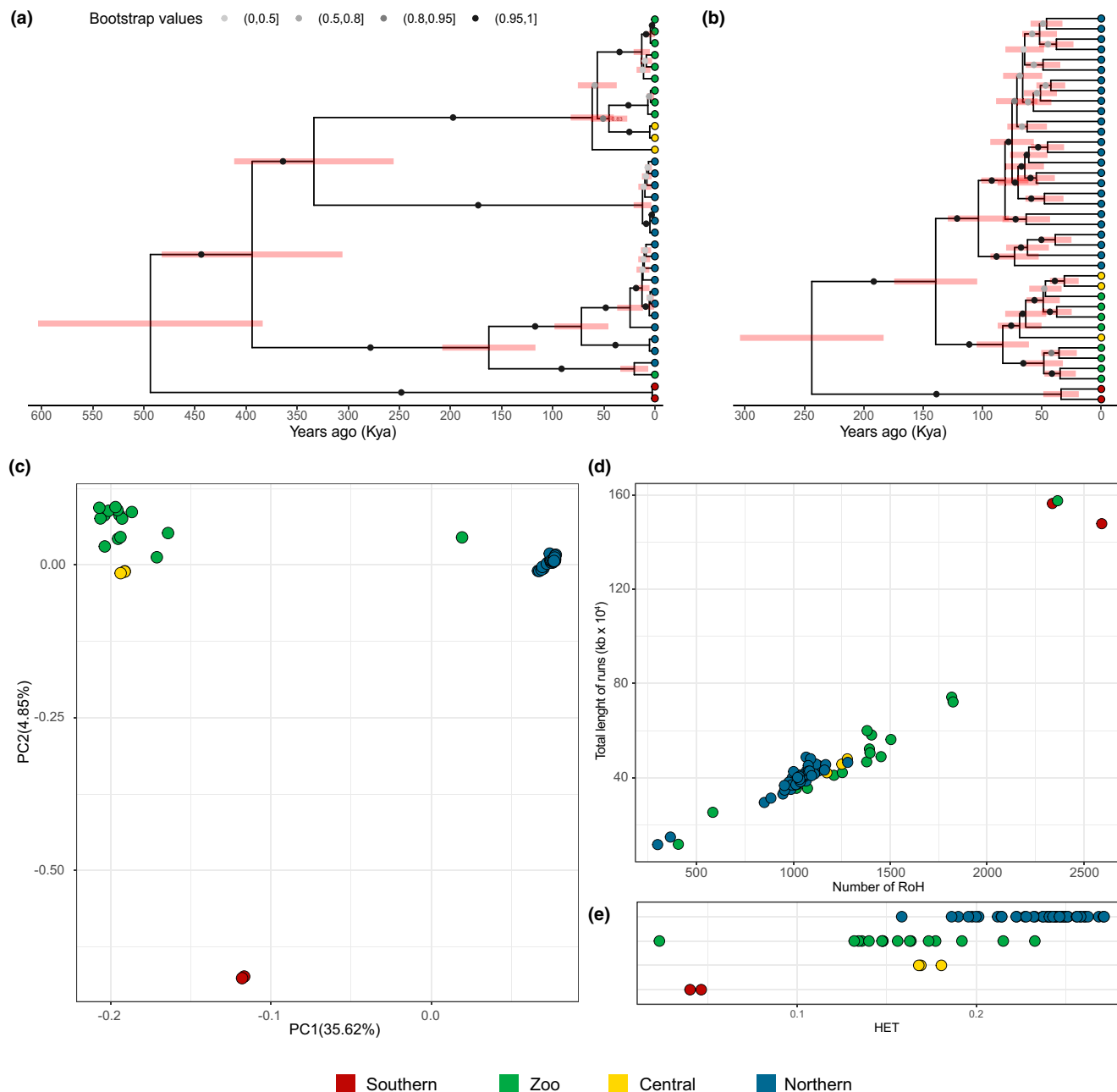


FIGURE 2 Population structure of *Theropithecus gelada*. Colours referred to gelada taxa as in Figure 1, with the addition of samples from zoos, in green. (a) Whole mitogenomes and (b) Y chromosome sequence phylogenetic trees; the red bars refer to highest posterior density (see Table S2), and the grey dots at the nodes refer to bootstrap values as indicated in the legend in panel a. (c) Principal component analysis of autosomal SNPs. (d) Intra-individual variation, reported for each sample as number of runs of homozygosity (RoH, x-axis) and total length of RoH (y-axis). (e) Autosomal SNPs proportion of observed heterozygous sites (HET) in investigated samples partitioned by population.

3.2 | *T. gelada* population structure and genomic variation

For the first time, the whole genome of individuals sampled from the gelada population located south of the Rift Valley (*T. g. arsi*) was investigated. The PCA results confirm the presence of structure within the species. The newly analysed Southern population of *T. gelada* is distinct from the two previously reported populations (Northern and Central), as suggested by autosomal STRs and mitochondrial

DNA (Shotake et al., 2016; Trede et al., 2020; Zinner et al., 2018; Figure 2a–c). The Northern population is separated from the other two along principal component 1 (PC1) which summarises 35.62% of the total diversity, while the Southern population separates from the others along PC2 which comprises 4.85% of the variation. As shown by Chiou et al. (2022), individuals from zoos belong to the Central population as our sample from the zoo in Stuttgart (TGLZ1).

Compared to other geladas, Southern individuals are clear outliers for the proportion of observed heterozygous sites, total length

and number of RoH. The only exception was SAMN20949854, a zoo sample that showed a lower heterozygosity, a greater number and a longer total length of RoH than other zoo samples (Figure 2d,e) whose coverage at 1.29x is also much lower than the rest of the samples here analysed (mean 14.5x; Table S1). We also tested the relationship between individuals up to the third degree and we noted that three pairs of geladas individuals from the Northern (Sankaber) population have a first-degree relationship; nine pairs have a second-degree relationship, including the two from the Southern population (sampled in different locations); and another nine pairs have a third-degree relationship, including two involving pairs of captive animals, one being TGLZ1 (Figure S2). It is worth noticing here that for these estimates the allele frequencies consider all the individuals here analysed. As such this might result in artificial closer affinities if the source populations have substantial lower diversity, as possibly the case for the Southern population.

3.3 | Demographic history of gelada populations

The reconstruction of demographic changes across time was performed using the PSMC software (Li & Durbin, 2011; Figure 3a). Results indicate that all three gelada populations experienced decreasing population size and/or connectivity over the last half a million years, differently from what was experienced by two geographically close Ethiopian Papionins, *Papio anubis* and *P. hamadryas* (Figure S3). Across geladas, the magnitude of this reduction was more pronounced for the Southern population (Chiou et al., 2022). The size of the three populations remained small over the last 50,000 years.

In order to have an indication of the timing of the most recent events of gene flow between different gelada populations, we

reconstructed the demographic curves of artificial F1 hybrids of chromosome X through hPSMC (Cahill et al., 2016; Figure 3). In accordance with coalescent theory the absence of coalescent events is indicative of extremely large populations, which results in the curves exponentially increasing at the time of the latest MRCA. The curves for the hybrids between gelada populations show this pattern at 10–20 kya, suggesting this was the time of last genetic exchange between groups. We observed the same results when the hPSMC analyses were replicated making use of the reads mapped to the *Mmul_10* reference genome (Figure S4b).

We then used D-statistics to further investigate patterns of gene flow across gelada populations by alternatively placing genomes of geladas from the three populations in position H1, H2 or H3. A gene-flow signal between Southern and Central geladas was highlighted when compared to the two Northern populations of Sankaber and Chenek (Figure S4a), and the two Northern populations showed a signal of genetic exchange with the Central than with the Southern population ($D=0.1199$, $Z=16.952$ and $D=0.1206$, $Z=16.627$) (Figure S4a). We also noted that the two Northern groups (Sankaber and Chenek) exhibited similar genetic exchange patterns with the Central and the Southern populations ($D=0.0021$, $Z=0.613$ and $D=0.0012$, $Z=0.301$ respectively); as such, we merged individuals from two Northern groups into a single population for subsequent analyses (Figure S4a).

3.4 | *Theropithecus*–*Papio* hybridisation

Following previous suggestions of possible *Papio* and *Theropithecus* hybrids based on molecular data and observations in the field (Dunbar & Dunbar, 1974; Walker et al., 2019), we tested for patterns

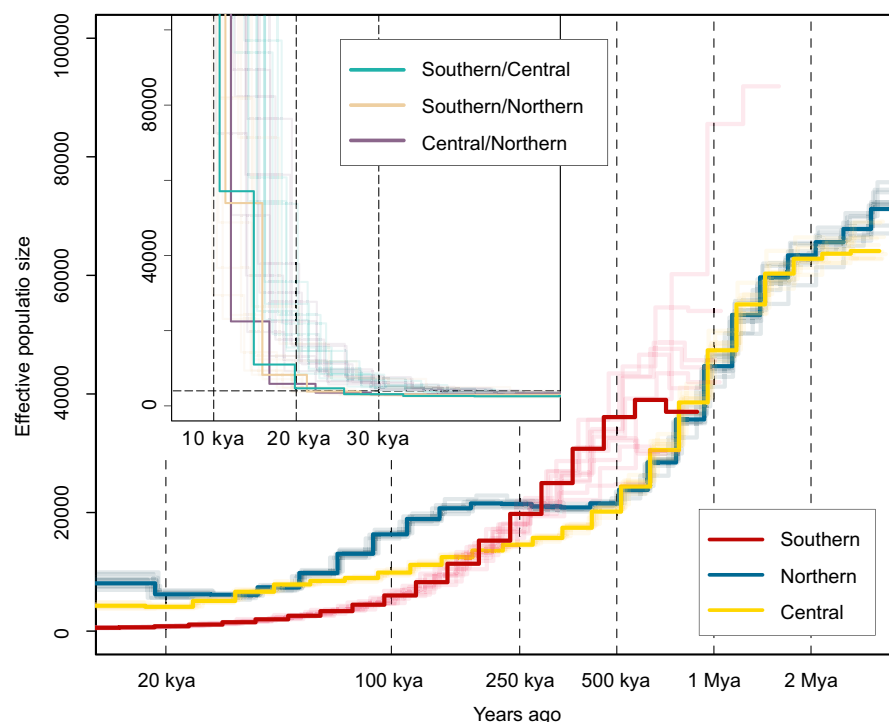


FIGURE 3 Demographic history of gelada populations. Changes in effective population size across time estimated using PSMC. SAMN20949878, SAMN20949865 and A28 genomes were analysed for Northern Central and Southern geladas respectively. X-axis on logarithmic scale. Colours as in Figure 1. Inset: hPSMC coalescence curves of artificial X chromosome hybrids of the same three samples, mapped on Tgel_1.0 reference genome.

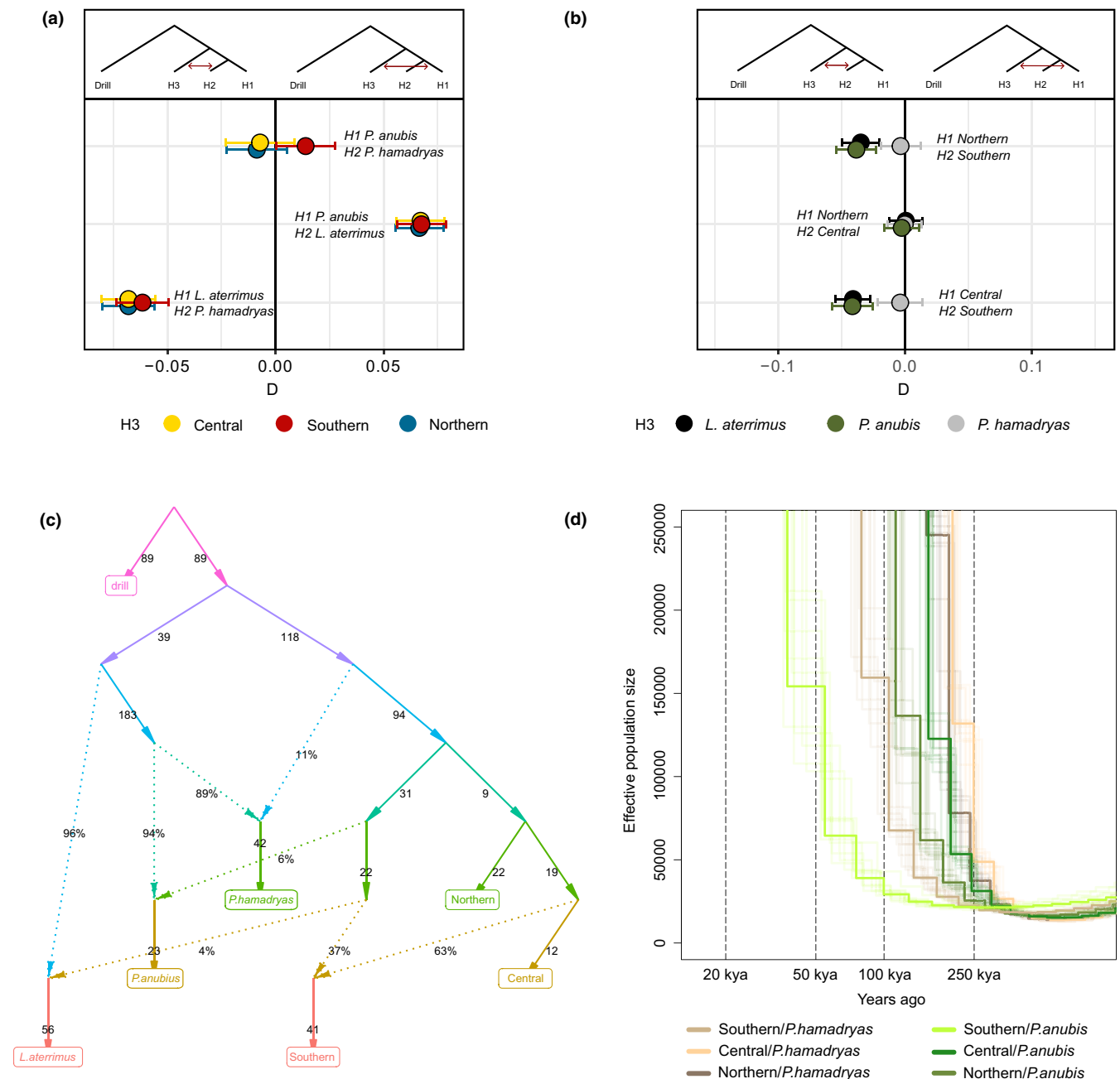


FIGURE 4 Gelada-baboons hybridisation events. (a) ABBA-BABA tests comparing *Theropithecus gelada* populations (H3) and baboons sp. (as H1 and H2); (b) ABBABABA test comparing baboons sp./*Lophocebus aerrimus* (as H3) and *T. gelada* populations (as H1 and H2); bars show the extent of three standard deviations; (c) admixture graphs (built using the R command `find_graphs`), reconstructing relationships and admixture events between Papionines species; dotted arrows indicate admixture, while associated numbers report the amount of each contribution; continuous arrows indicate linear descendants, while associated numbers are indicative of the amount of drift experienced along lineages, given in f-statistics units multiplied by 1000; (d) coalescence of artificial hybrids *Papio-Theropithecus* X chromosomes (hPSMC analysis); the gelada samples used are the same in Figure 3; samples for this analyses were mapped on Mmul_10 reference genome); x-axis on logarithmic scale.

of shared alleles across the two genera using the *D*-statistics. We initially tested for any gene-flow signal between gelada and *Papio* genomes when compared to *Lophocebus*. Being sister clades, in the absence of gene flow, *Papio* and *Lophocebus* should be similarly related to *Theropithecus* (Perelman et al., 2011). Our results

highlighted that all *Papio* genomes exhibited more shared alleles with geladas than *Lophocebus*, suggesting contacts between the two genera (Figure 4a). In addition, *Papio* species showed no difference in their degree of shared alleles with different gelada populations and vice versa, except for *P. anubis* and Southern geladas among

which was highlighted a signal of gene-flow (D(Drill, Southern, *P. hamadryas*, *P. anubis*) Z-score=3.038; D(Drill, *P. anubis*, Southern, Central) Z-score=-7.916; D(Drill, *P. anubis*; Southern, Northern) Z-score=-7.410; Figure 4a,b). Overall, these results suggest that gelada populations differ in the amount of between species introgression that occurred in their evolutionary history.

We investigated the evolutionary relationships among *T. gelada* populations and *Papio* species (and *Lophocebus*) by means of the approach implemented in *find_graphs* (Figure S5). The topology with the best likelihood score, the greatest number of admixture events and no branches with zero drift estimates is presented in Figure 4. The graph underlined the occurrence of an ancient contribution of 11% from the *Theropithecus* branch to the lineage leading to *P. hamadryas* and subsequent smaller contributions to *P. anubis* (6%) and *L. aterritus* (4%) from a population which also contributed, together with a pre-Central group, to Southern geladas (Figure 4c).

The generation of *Theropithecus*-*Papio* hybrid X-chromosomes and their analysis via *hPSMC* provided the chronological context for the *Papio*-gelada contacts suggested by the *D*-statistics and highlighted by *find_graphs*. All *P. hamadryas*/*T. gelada* hybrids generated curves older than 100 kya. Similar curves were generated by *P. anubis*/*T. gelada* hybrids except for the hybrids involving *P. anubis* and Southern geladas which were more recent (younger than 50 kya; Figure 4d). We observed the same results when the *hPSMC* analyses were replicated making use of the reads mapped to the *Tgel_1.0* reference genome (Figure S6).

3.5 | Copy number variants

We investigated the pattern of copy-number variants (CNVs) across the three gelada populations by leveraging the population-based genomic data available and newly generated here. A total of 656 variable regions with more than two genomic copies were identified across the 69 available gelada individuals (Table S4). The size of the longest expanded regions was up to 5.5 Mb, the vast majority being below 1 Mb (Figure S7a). Notably, 45% are singletons, appearing only once in the whole gelada dataset (Figure S7b). 243 of the non-singletons were found only in one of the three populations (Figure 5a). We considered that local positive selection can drive the increase in the frequency of specific variants in a given population and therefore focused our attention on the set of population-specific CNVs that were present in no less than 25% of individuals in a population (5 in Central and 13 in Northern populations). As only two individuals were analysed for Southern geladas, we considered all the regions uniquely expanded in this population and present in both individuals. In this way, we identified 20, 17 and 10 regions in the Northern, Central and Southern populations, which contained 19, 9 and 6 genes respectively. No significant enrichment in KEGG descriptors was found for the 34 genes across the three populations. We manually inspected the list of genes and noted, among the others, the occurrence of one of the two pancreatic amylase genes (LOC112607926). The two Southern geladas genomes hosted a different number of amylase duplications: three copies were inferred for A28, while four were estimated in A10, the rest of the

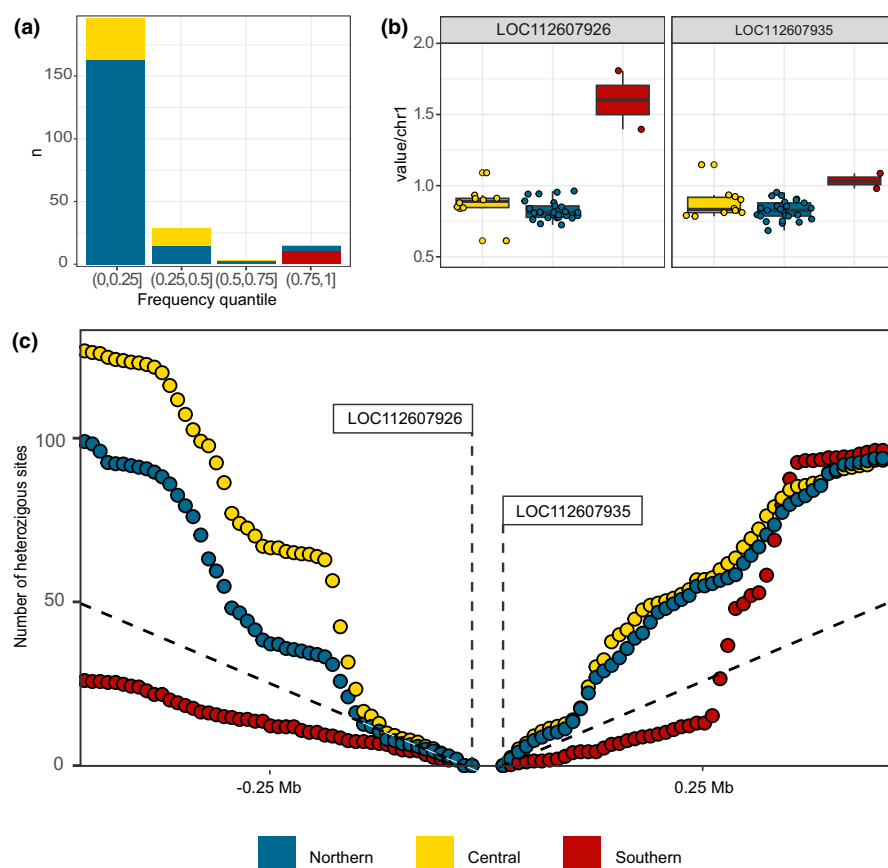


FIGURE 5 CNVs and amylase gene copy number. (a) Population-specific CNVs divided according to their frequency in each population. (b) Normalised genomic coverage over the two loci annotated as pancreatic amylase in geladas. (c) Cumulative number of heterozygous positions in 10-kb windows, starting from the beginning/end of the two geladas pancreatic amylase gene average and extended for 0.5 Mb. Reported is the average across individuals for the indicated population, each individual estimate normalised for the average number of heterozygous positions estimated on chromosome 1 in each individual (see Section 2); dashed lines refer to the expected cumulative curve for 10-kb windows each displaying the average number of heterozygous positions along chromosome 1.

geladas presenting only two (Figure 5b). The extent of the duplication was approximately 50kb in both individuals (spanning positions 131,716,000–131,771,999 and 131,720,000–131,771,999 in samples A28 and A10 respectively; Figure S7c). Given the reliance of geladas on underground food in the dry season, we decided to further explore the pattern of variation associated with the amylase sequences duplicated in Southern geladas (LOC112607926).

Considering a 10-Mb region surrounding the duplicated amylase locus, Southern geladas showed a decrease in diversity that extended over a chromosome portion much larger than other populations, up to Mbs in size (Figure 5c and Figure S7d). The physical extent of the reduced diversity as the result of positive selection is shaped by recombination events, which are scaled by the number of generations since the selective pressure started. The observed pattern in Southern geladas is compatible with positive selection on the duplicated amylase gene being relatively recent (Sabeti et al., 2002, 2007).

3.6 | Variation in underground food consumption and rainfall

We tested for association between rainfall and UF consumption using the full UF dataset of the five locations investigated here (Sankaber and Gich, Northern; Guassa and Wollo, Central; Indetu, Southern; Fashing et al., 2014; Jarvey et al., 2018; Kifle & Bekele, 2021; Woldegeorgis & Bekele, 2015) and data on monthly rainfall in the same locations (Fick & Hijmans, 2017; Harris et al., 2020; Jarvey et al., 2018). UF consumption was significantly associated with four tested rainfall estimates (Jarvey et al., 2018): same month total rainfall, $r_s = -.43126$, $p(2\text{-tailed}) = .0003$, monthly rainfall during the previous month, $r_s = -.349$, $p(2\text{-tailed}) = .00408$, cumulative rainfall during previous 2 months, $r_s = -.34862$, $p(2\text{-tailed}) = .00412$ and cumulative rainfall during previous 3 months, $r_s = -.2801$, $p(2\text{-tailed}) = .02274$ (Jarvey et al., 2018). Similar patterns are observed when the cumulative estimates include the month when UF consumption was considered (Figure S8a). We verified if any region affected in a substantial way the results by repeating the analysis removing one dataset per time. The removal of the data from Indetu, but not of other datasets, substantially improves the correlation between monthly rainfall and UF usage (Figure S8b).

The five datasets were collected in different years, but all included the same 6-month period (August–January). UF consumption during this period across the five locations is significantly different (One-Way ANOVA, $F_{4,25} = 3.82243$, $p = .014$), pairwise comparisons being significant for the Indetu–Sankaber pair (post-hoc Tukey test, $Q = 5.13$, $p = .01$; Figure S8c). During this 6-month period, the average UF contribution to diet in Indetu (31.25%) was 50%–300% higher than in the other four regions (range 8.5%–22.21%). Using a moving window of seven consecutive months (the longest consecutive period available for Indetu), the distributions of estimated averages for Northern and Central overlap, but the mean value across seven continuous months at 33% in Southern is outside the

variation reported for the other two taxa (Abu et al., 2018) (the maximum values in the four populations being 15%–26%; Figure S8c; Abu et al., 2018). The main differences between Indetu and the other locations appear to be not in the maximum monthly UF consumption (46.6% in Indetu, 49.7% in Sankaber, 36.7% in Gich, 28.1% in Guassa, 57.7% in Wollo) but in this being consistently above 20% in Indetu (over the 9 months of available data) but not so in the other locations (3 months in Sankaber, 5 in Gich, 6 in Guassa and 7 in Wollo; Figure S8c). We tested for differences in rainfall across the five regions but no significant differences were found for any of the four rainfall estimates related to the overlapping 6-month period (August–January) (one-way ANOVA, monthly rainfall during relevant month, $F_{4,25} = 0.03729$, $p = .99716$; monthly rainfall during the previous month, $F_{4,25} = 0.4303$, $p = .7854$; cumulative rainfall during previous 2 months, $F_{4,25} = 0.7754$, $p = .5516$; cumulative rainfall during previous 3 months, $F_{4,25} = 1.3515$, $p = .2789$). The same was also the case for cumulative estimates including the month when UF consumption was considered (data not shown).

3.7 | Paleoclimate and gelada spatio-temporal distribution

We investigated the eco-climatic similarity of the regions currently occupied by geladas using climatic descriptors of the regions where they occur. Our analysis did not highlight significant climatic niche overlap among the three gelada populations (Figure 6 and Figure S9). Nonetheless, our results suggest the Southern population niche resembles more that of the Central population than the one of the Northern geladas. The two in fact partially overlap in their ecological spaces but the habitat of the Southern population is characterised by lower temperature values during the coldest and driest months (bio9 and bio11; Figure 6). Our results also indicate that the Northern population is characterised by the highest climatic variability. In contrast to the other two populations, the density of the Northern population extends along the PC1 axis towards higher bio15, bio18 and bio14 values suggesting that this population is associated with wetter habitats with overall more precipitation (Figure 6).

SDMs achieved excellent performance with a weighted mean AUC = 0.96 (sd = 0.06). SDM predictions suggested a high climatic-driven variability in habitat suitabilities for the three gelada populations in the past. Northern and Central populations show similar behaviour in terms of climatic tolerance and adaptation. In contrast, we observed a general lower habitat suitability for the area where the Southern population actually occurs. In addition, we noted a drastic reduction in habit suitabilities for geladas during the Last Glacial Maximum (LGM; 24–18 kyr; Figure 7; Supplementary Material 1). Despite an overall post-LGM improvement, the habitat suitability of the region occupied by the Southern population resulted consistently lower than that of the regions occupied by the Central and Northern populations (Figure S10; Supplementary Material 1).

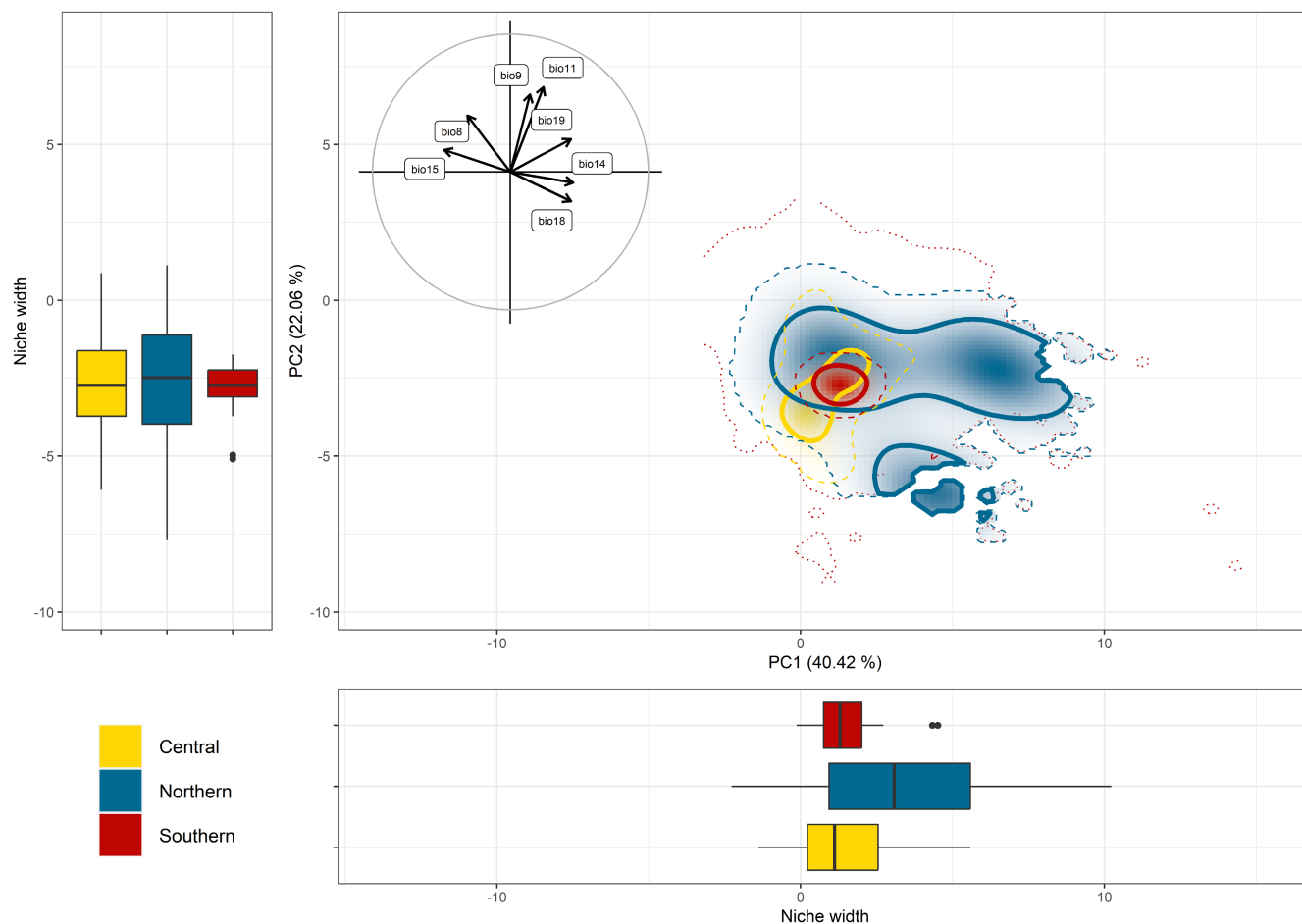


FIGURE 6 Gelada population niches along the first two principal component analysis (PCA) axes (PC1 and PC2) of bioclimatic variables recorded at occurrence sites and in the background environment within the study area. Red: niche calculated from central population sites; yellow: niche calculated from southern population sites; blue: niche calculated from exclusive northern population sites. Colour shading shows the density of the occurrences of the species by cell, with the solid and dashed contour lines illustrating, respectively, 50% and 90% of the available environment.

4 | DISCUSSION

4.1 | Population dynamics and climate changes

The role played by climate variation on the evolutionary history of extant and extinct species has gathered centre stage in the discussion about major drivers of evolution (Carotenuto et al., 2016; Nogués-Bravo et al., 2010). Long and short-term wet-dry cycles have been investigated in relation to long-distance dispersals and isolation, events of gene-flow and hybridisation, drift and local adaptation in human and non-human primates (Chala et al., 2019; Stewart et al., 2020; Timmermann et al., 2022; Timmermann & Friedrich, 2016). Dietary and environmental specialisations (Chiou et al., 2022), together with a low fertility rate (one infant every 2–3 years; Dunbar, 1980) make gelada populations more vulnerable to eco-climatic changes, past and present. As such the reconstruction of how climate changed in the past can provide the context for interpreting the demographic events inferred from the analysis of genomic variation in gelada populations. The analysis

of autosomal microsatellites and mitochondrial DNA sequences in several *T. gelada* individuals sampled in the wild has previously suggested the subdivision of the species into three main populations, generally referred to as Northern, Central and Southern in accordance with their geographical distribution along Ethiopian highlands (Gippoliti, 2010; Shotake et al., 2016; Trede et al., 2020; Zinner et al., 2018). Our genomic analysis, extended to include data from gelada individuals collected from the Southern population, confirmed the proposed tripartition of this species. Full mitogenomes and Y chromosome sequences were similarly in agreement with the three-population structure, with further subdivisions hinted by mtDNA and genomic data, but not Y chromosome sequences (Figure 2a,b). Male lineages from different Northern subpopulations did not cluster according to their sampling location (unfortunately, having only one subpopulation for Central and Southern gelada, we were not able to explore this aspect in these populations) and lineages from different regions were related to each other. Such lack of structure might reflect the predominantly male-mediated intergroup migration observed in the wild

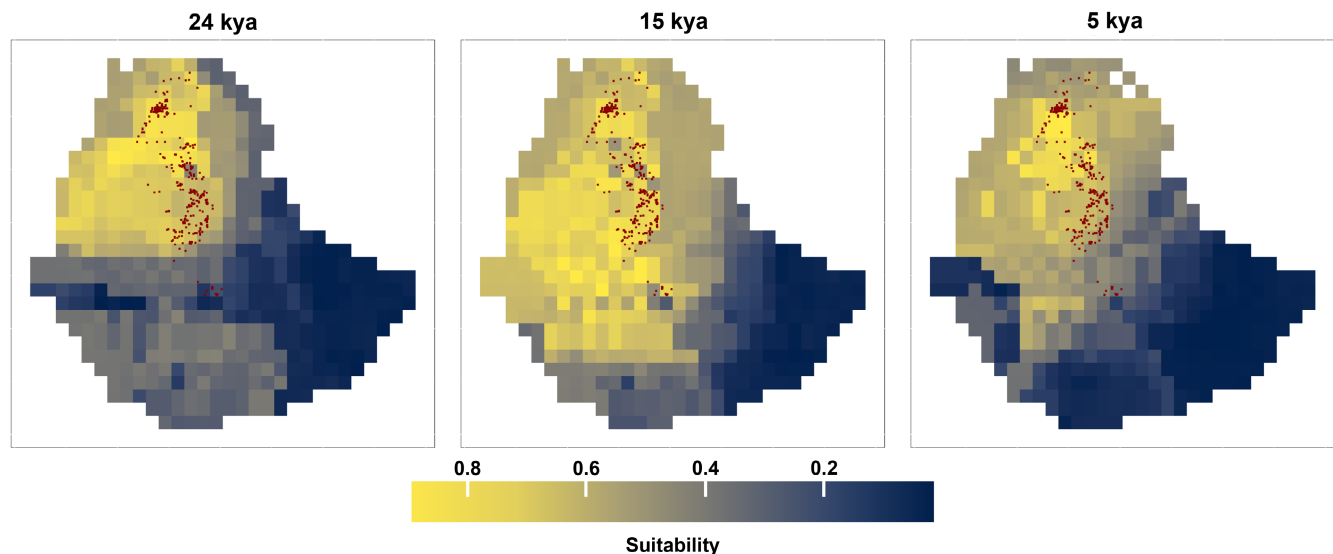


FIGURE 7 Predicted habitat suitability maps of *Theropithecus gelada* in Ethiopia at three different past time frames. Higher values indicate the areas that were more favourable for gelada habitation. Red dots represent the current geographical location of the gelada sites.

(Dunbar, 1980; le Roux et al., 2011). The different demographic dynamics affecting males and females might also explain the significant differences in the amount of variation recovered using the mtDNA and the Y chromosome as summarised by their different TMRCA, the Y chromosome being two times younger than the mitogenome (0.24 Mya vs. 0.49 Mya). Similar differences in mitochondrial and Y chromosome TMRCA have been noted previously in gorilla/orangutans but not in chimpanzees (Hallast et al., 2016). Such differences have been explained by non-mutually excluding scenarios related to the mating strategy implemented by males: (a) a substantial skew in male reproductive success, which would result in a reduction in Y chromosome diversity; (b) a reduced sperm competition, resulting in lower sperm production and an associated lower Y-related mutation rate. A lower mutation rate has not been reported in baboons, but male-skewed reproductive success has been described (Fischer et al., 2019; Wu et al., 2020). It is therefore possible that the skewed sampling process associated with male variation reproductive success might be the underlying explanatory mechanism for the observed Y/mtDNA TMRCA in geladas. Relative differences in age are also dependent on the accuracy of mutation rates for different loci (Hallast et al., 2016). Further investigations into other species might provide useful insights on this topic.

The lack of shared Y chromosome and mtDNA lineages across the three gelada populations suggests a relatively deep subdivision among these groups, the mtDNA/Y chromosome TMRCA between Southern and Central-Northern gelada dating to 500–150 kya. Gene-flow among populations appears to have been interrupted only in the last 20 kya but population contractions had been in place much earlier than that (Figure 3). The lack of more recent coalescent events for uniparental markers could be due to sampling as well as drift, the latter expected to be more pronounced for haploid systems. The occurrence of deep

mitochondrial and Y chromosome branches in Southern geladas is compatible with a model suggesting early colonisation of the region followed by periods of isolation and population contraction (Supplementary Material 1; Figure S10). Interestingly, the most recent period of isolation suggested by genomic data appears to parallel changes in the climatic-driven habitat suitability of *T. gelada*, with gene-flow being interrupted since LGM (Figures 7; 3). A dynamic, metapopulation model is suggested by admixture trees, PSMC plots and paleoclimate reconstructions, all depicting a scenario where events of expansion, gene-flow, isolation and possibly local extinction were probably not uncommon (Supplementary Material 1). The extreme fluctuations in habitat suitability are also compatible with the continuous reduction in effective population size estimated for *T. gelada* (Figure S10, Supplementary Material 1; Figure 3), resulting from decreasing population size and/or connectivity. The more extreme reduction in diversity and increase in RoH reported for Southern geladas parallel the more severe change in effective population size highlighted over recent times by PSMC and follows the lower degree of habitat suitability experienced over time by the Southern population (Figures 2d,e, 3, 7; Figure S10, Supplementary Material 1).

4.2 | Hybridisation

Our extensive genomic comparison provided support for previously suggested gene-flow between *Theropithecus* and *Papio*, their last common ancestor possibly dating to 5 Ma, close to the human-chimpanzee split (Liedigk et al., 2014). Hybrids have been potentially identified in the wild (Dunbar & Dunbar, 1974) and observed in captivity (Jolly et al., 1997; Markarjan et al., 1974). The phenotypes of the captive and wild hybrids were different, possibly reflecting gelada or *Papio* backcrossing (Jolly et al., 1997). Captive F1 showed

intermediate or more gelada-like features, exceeding both parental populations for body size (Jolly et al., 1997; Markarjan et al., 1974). Our data suggested ancient gelada contributions to *Papio* species present in Ethiopia (*P. hamadryas* and *P. anubis*). Interestingly, the time of the last contact of *Papio anubis* with Southern geladas aligned with a period when the predicted Southern geladas suitable habitat area was increasing (Figure S10, Supplementary Material 1). Similar changes have occurred multiple times in the past and might have facilitated also the more ancient contacts of *T. gelada* with *P. hamadryas*.

The occurrence of captive fertile F1 (Markarjan et al., 1974) but the lack of genomic evidence for recent gene-flow could be explained by the hybrid F1 being generally disfavoured in the wild and therefore less likely to survive and reproduce. Jolly and colleagues (Jolly et al., 1997) proposed two possible explanations of why hybrids might have a lower fitness in the wild, both based on their larger appearance. Both considered the additional energy requirements associated with larger body size: increased nutritional requirements but lower-crowned teeth (as observed in the hybrids in the wild; Jolly et al., 1997) might drive early loss of dental function and inability to fulfil their nutritional needs, and the increased maternal burden associated with the raising of a larger foetus and the nursing of a bigger infant (Jolly et al., 1997).

The antiquity of the introgression events probably provided enough time for natural selection to remove incompatible or deleterious gelada alleles from the genome of *Papio* (and *Lophocebus*), as well as increase the frequency of favourable variants, in a scenario not too dissimilar from what was observed in humans after the hybridisation with Neanderthals and Denisovans (Tobler et al., 2023). The recent release of a large *Papio* genomic dataset is expected to enable a more detailed investigation of the genomic distribution of gelada ancestry and their possible phenotypic impact (Groh & Coop, 2024; Sørensen et al., 2023).

4.3 | Climate, diet and local adaptations

The unique dietary profile of geladas dominated by grass makes the variation across populations in UF usage of interest (Figure S8; Jarvey et al., 2018). Grass is the staple of choice of geladas, its contribution to diet being shaped by its availability ultimately influenced by rainfall. UF operates as fallback food, its usage strongly driven by grass availability. While the link between grass consumption and rainfall is strong, the one between dietary UF and rainfall is present but weaker (Jarvey et al., 2018 and our results). Other factors probably play a role in determining UF usage (Abu et al., 2018). All this considered, despite no variation in rainfall, geladas of the Southern population of Indetu appear to consume more UF than other gelada groups. Interestingly Southern geladas harbour more copies of LOC112607926, one of the two genes annotated as pancreatic amylase in the gelada reference genome (Figure 5), a genomic feature observed in dogs, pigs and human populations with a diet enriched in starchy food (Jarvey et al., 2018; Pajic et al., 2019; Perry et al., 2007).

The reduction in diversity around the duplicated locus in Southern gelada is indicative of positive selection, while the chromosomal extent of this pattern suggests this being a relatively recent event. For comparison, the reduction of diversity around the variants associated with lactase persistence in humans is similarly observed over more than 1 Mb (Bersaglieri et al., 2004; Joslin et al., 2020). Considering that selection on the persistence of LCT expression into adulthood has been dated to no more than a few thousand years ago (considering a generation time in humans of 28 years, these correspond to a few hundred generations; Burger et al., 2020; Evershed et al., 2022), it is reasonable to consider the pattern observed in Southern gelada being generated over a number of generations around the same order of magnitude (at least hundreds).

During the past 30 kya, southern Ethiopia has undergone strong climatic changes, from dry and relatively cold during the Last Glacial Maximum (LGM, 25–18 kya) to the African Humid Period (AHP, 15–5 kya; Casas-Gallego et al., 2023; Fischer et al., 2021; Foerster et al., 2012). Our past SDM predictions perfectly fit the Ethiopian climatic trends suggesting that gelada spatial distribution was affected by these intense climatic fluctuations (Figure 7). Particularly, we show that during the LGM the Southern geladas extremely suffered cold and dry conditions (Figure 7; Supplementary Material 1; Figure S10) whereas the impact was limited for the other two populations. We propose that increased UF consumption and amylase copy number are related in Southern geladas. The ecological context experienced over time by this population might have driven positive selection on the amylase duplication in the last few thousand years, compatible with the physical extent of the reduction of diversity around this locus. Increased UF consumption by geladas has been observed when grass/rain is scarce and metabolic adaptation enabling more efficient processing of starchy food is expected to provide an evolutionary advantage in such situations (Jarvey et al., 2018; Pajic et al., 2019; Perry et al., 2007). While past climate might have provided the context for positively selecting a higher number of copies of the amylase gene in Southern geladas, other dynamics might have contributed over time too. The observation that increased UF consumption in Indetu occurred despite no differences in rainfall might reflect behavioural differences in food preference across gelada populations. It has been also reported that gelada UF consumption in Indetu in the wet season is slightly higher than during the dry season (seasonal mean 38% and 33% respectively; Abu et al., 2018). A general, behavioural-related, increase in UF consumption could have reinforced the climate-related selective pressure on salivary amylase copy-number, generating a co-evolutionary scenario not too dissimilar from the gene–culture interactions reported in humans for lactase persistence (Evershed et al., 2022). It remains to be tested if in Indetu the use of UF is less related to the amount of grass available and if local behavioural traits contribute to the increase in UF consumption (Whiten et al., 1999). Interestingly, the removal of data from Indetu improves the correlation between UF consumption and rainfall in the dataset we tested, suggesting that other variables, including anthropogenic, affect dietary UF contribution to diet in Indetu (Abu et al., 2018).

The Ethiopian Highlands (part of the Eastern Afromontane Biodiversity Hotspot; Mittermeier et al., 2011) can be divided into the main north-western massif (western plateau) and the smaller south-eastern plateau, the two separated by the East African Great Rift Valley (Yalden et al., 1996). The highland plateaus belong to the same ecozone (Afroalpine and subafroalpine vegetation) although, some variation in the taxonomic composition of both plant and animal communities as a result of biogeographical patterns across gelada locations is present (Vivero et al., 2006). Currently, the lack of a systematic investigation in plant composition across the plateaus and locations of gelada occurrence prevents any further analysis in how these affect food consumption. Nevertheless, it is worth noticing that gelada on the Eastern plateaus live at slightly lower altitudes than those on the western plateau (1600–2300m vs. above 2300; Abu et al., 2018; Belay & Shotake, 1998), which might further affect the set of species available to geladas. However, it will be necessary to collect more information on what plant species are present and the extent of their occurrence across sites to test how these aspects shape what geladas prefer and ultimately consume.

We note that the two Southern Individuals here whole genome sequenced were not collected in Indetu and it remains to be shown if increased copy number/increased UF consumption are features shared across all Southern geladas. More ecological and behavioural data, as well as a more extensive characterisation of amylase copy-number variation and their expression in Southern geladas, are necessary to test the association between UF consumption, grass availability and amylase genes in this population.

We note that an increase in the copy number of pancreatic amylase is expected to increase not only the expression in the pancreas but also in the salivary glands, as pancreatic amylase is also expressed in salivary glands (Axelsson et al., 2013; Fagerberg et al., 2014; Han et al., 2022; Mau et al., 2010; Pajic et al., 2019; Samuelson et al., 1996; Torres et al., 2018). Selective pressure associated with an increase in starch consumption have been shown to favour the expansion of amylase genes, salivary and/or pancreatic (Pajic et al., 2019). The related increase in amylase production, both in the pancreas and the salivary glands, results in an increased capacity of processing starch both in the mouth and the digestive system (Axelsson et al., 2013; Mau et al., 2010; Pajic et al., 2019; Perry et al., 2007; Samuelson et al., 1996), which might be the case also for Southern geladas.

4.4 | Implications for the taxonomy and conservation of *T. gelada*

The genomic analyses here presented support that presently the genus *Theropithecus* may be subdivided into three evolutionary units that have been taxonomically designated as *Theropithecus gelada gelada* (Rüppell, 1835), *Theropithecus gelada obscurus* Heuglin, 1863 respectively in the northern and central areas of the Ethiopian plateau, and, in the south-eastern mountain range, *Theropithecus gelada arsi*, so agreeing with most recent literature (Gippoliti, 2010; Shotake et al., 2016; Trede et al., 2020; Zinner et al., 2018). This taxonomic

account represents a solid base for conservation policies that aim to protect the current diversity of the genus, but, as often the case, must not be considered definitive. In particular, it remains to be determined whether the three identified evolutionary units should be treated as having equal rank or if the Arsi (Southern) gelada should be considered a distinct species. This consideration is based on the potentially older divergence of the Arsi gelada from the two subspecies, *gelada* (Northern) and *obscurus* (Central), suggested by deep coalescent events of uniparental markers and is consistent with the Rift Valley's well-known role as a barrier for mountain animal specialists in Ethiopia (Lavrenchenko & Bekele, 2017). Additionally, the ecological challenges of the Arsi habitat, supported by food-related adaptations reported for the amylase gene and the specific present and past climatic niche of the Arsi geladas, further support treating the Arsi population as a distinct species. We highlight that other standard taxonomic databases, such as cranial morphology, have been scarcely investigated so far and could be critical to further refine our taxonomic knowledge.

In this regard, we also note that the Northernmost gelada populations have been recently confirmed in the Tigray, whose genetic and morphological affinity to other gelada populations is still unclear (Girmay & Dati, 2020; Girmay & Tesfay, 2020; Haileselasie et al., 2023). Considering the geographical features of the region, it is not impossible that the geladas from Tigray will turn out to be closer to *T. g. obscurus* (Central) than *T. g. gelada* (Northern), making the geographical designation of these subspecies as Central and Northern incorrect and instead supporting an alternative nomenclature as Western (*gelada*) and Eastern (*obscurus*) geladas (Gippoliti, 2010). Future characterisation of these groups will clarify this interesting geographical and taxonomic aspect of the species.

As a species *Theropithecus gelada* is considered Least Concern by IUCN (Gippoliti et al., 2019). Irrespective of the taxonomic rank accorded, *T. gelada arsi*, with between 2000 and 3000 surviving individuals estimated (De Jong et al., 2023), is a conservation priority despite the lack of a formal description that has delayed a formal recognition by IUCN (Gippoliti, 2022). The Arsi gelada may be considered Endangered according to IUCN criteria – a formal assessment has just been published (De Jong et al., 2023). The Galama Ridge, which offers protection to the Arsi geladas, suffers from intense anthropogenic pressure and no protected areas exist in the region. The remaining mosaic forests are threatened by intensive deforestation and agricultural ploughing. Montane habitats suffer from burning during the dry season and overgrazing during the wet season (Kostin et al., 2019). This taxon may therefore serve as a much-needed flagship species attracting naturalistic tourism to their narrow region of occupancy along the gorges of the Webi Shebeli and Robi rivers. Obviously, revenue from tourism must be shared with local communities to lessen anthropogenic impacts on ecosystems.

5 | CONCLUSIONS

By integrating genomic, climatic and dietary data we have generated a detailed picture of the evolutionary history of *Theropithecus gelada*.

The inclusion of more genomic data, from living individuals and museum specimens (Gippoliti, 2010), is expected to refine our results by providing a more detailed description of the degree of population structure and gene flow across populations that may be critical for long-term conservation planning. The integration of morphological data and the inclusion of functional analyses will contribute additional insights on the degree of variation within the species and the evolutionary significance of such diversity. Reconstructing the history of this species is also expected to unlock to what extent geladas will be able to face the challenges related to ongoing and future climatic changes (Ahmed et al., 2023). Indeed, a formal assessment of this population has just been published (De Jong et al., 2023) that suggests the Arsi gelada should be classified as Endangered according to IUCN criteria.

AUTHOR CONTRIBUTIONS

MC and CC designed the study, with the contribution of SG. CC provided founding, supervised the data analysis and coordinated the project. TS and GB collected the southern gelada genomes. HI conducted the molecular evaluation of the southern gelada genomes and coordinated their sequencing. MC, GM, AM, GOG, JRT, JM and FM executed the analysis. MC, AM, GM, SG and CC wrote the manuscript, with contributions from all authors. All authors reviewed and edited the manuscript.

ACKNOWLEDGEMENTS

This work was supported by the funding of the Leakey Foundation to CC ('The Y chromosome evolutionary history of *Papio*'), the 'Departments of Excellence' program of the Italian Ministry for Education, University and Research (MIUR, 2018-2022). GM received a predoctoral fellowship from the 'Caixa' Foundation (LCF/BQ/DI22/11940014). FM was supported by 'Fondazione con il Sud' (2018-PDR-01136), by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.16-0030) and by the Italian Ministry of University and Research (2022P2ZESR). Open access publishing facilitated by Università degli Studi di Parma, as part of the Wiley - CRUI-CARE agreement.

This research benefits from: the HPC (High-Performance Computing) facility of the University of Parma, Italy; the PhD program in Biotecnologie e Bioscienze (University of Parma), the Cooperative Research of Primate Research Institute (Kyoto University), the Microbial Cellular and Molecular Biology Department of Addis Ababa University, the Ethiopian Wildlife Conservation Organization.

The authors would like to thank the European Association of Zoos and Aquaria Biobank (EAZA; Anna Mekarska), the Wilhelma Zoo Stuttgart, the Gaia Zoo Kerkrade and the EAZA Biobank hubs at the Leibniz-Institute for Zoo and Wildlife Research (Jörn Fickel, Anke Schmidt) and at the Copenhagen Zoo (Christina Hvilsom, Tanja Normann Gade, Rafaela Fiuza) for help and support in providing biological samples for genomic analysis. The authors would finally like to thank all the residents of the villages for their cooperation in the field during the 1978-1997 sampling visits.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

OPEN RESEARCH BADGES



This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at [[insert provided URL(s) on the Open Research Disclosure Form]].

DATA AVAILABILITY STATEMENT

The raw sequence reads here generated (samples A10, A28, TGLZ1 and LACZ1) are available under the ENA accession study number PRJEB72259 (accessions numbers from ERR12892801 to ERR12892810).

ETHICAL APPROVAL

The project has been revised and approved by the ethics committee for animal investigation of the University of Parma (OPBA, Committee for the Animal Wellbeing; PROT.N.1/CESA/2022).

CONSENT FOR PUBLICATION

All authors have provided consent for the publication of this paper.

BENEFIT-SHARING STATEMENT

In the framework of the Nagoya Protocol, our study encourages domestic-level benefit-sharing measures providing access to scientific information relevant to conservation and sustainable use of biological diversity, including biological inventories and taxonomic studies (section k); specifically helping to determine priority gelada populations for conservation measures. Hopefully, such conservation measures may in future include eco-tourism activities for distinctive geladas, such as the threatened small Arsi population, that may positively contribute to increase and diversify local economy (section l). This work also involved a local researcher (G.B.) from Addis Ababa University.

ORCID

Matteo Caldon <https://orcid.org/0000-0002-5108-2021>

Hiroo Imai <https://orcid.org/0000-0003-0729-0322>

Gonzalo Oteo Garcia <https://orcid.org/0000-0002-0957-4014>

Cristian Capelli <https://orcid.org/0000-0001-9348-9084>

REFERENCES

- Abu, K., Mekonnen, A., Bekele, A., & Fashing, P. J. (2018). Diet and activity patterns of Arsi geladas in low-elevation disturbed habitat south of the Rift Valley at Indetu, Ethiopia. *Primates; Journal of Primatology*, 59(2), 153-161. <https://doi.org/10.1007/s10329-017-0640-9>
- Ahmed, A. S., Chala, D., Kufa, C. A., Atickem, A., Bekele, A., Svenning, J. C., & Zinner, D. (2023). Potential changes in the extent of suitable habitats for geladas (*Theropithecus gelada*) in the Anthropocene. *BMC Ecology and Evolution*, 23(1), 65. <https://doi.org/10.1186/s12862-023-02173-3>

- Almarri, M. A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A. S., Chen, Y., Hurler, M. E., Tyler-Smith, C., & Xue, Y. (2020). Population structure, stratification, and introgression of human structural variation. *Cell*, 182(1), 189–199.e15. <https://doi.org/10.1016/j.cell.2020.05.024>
- Axelsson, E., Ratnakumar, A., Arendt, M. L., Maqbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, A., & Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, 495(7441), 360–364. <https://doi.org/10.1038/nature11837>
- Batra, S. S., Levy-Sakin, M., Robinson, J., Guillory, J., Durinck, S., Vilgalys, T. P., Kwok, P. Y., Cox, L. A., Seshagiri, S., Song, Y. S., & Wall, J. D. (2020). Accurate assembly of the olive baboon (*Papio anubis*) genome using long-read and hi-C data. *GigaScience*, 9(12), g12134. <https://doi.org/10.1093/gigascience/giaa134>
- Belay, G., & Shotake, T. (1998). Blood protein variation of a new population of gelada baboons (*Theopithecus gelada*), in the southern Rift Valley, Arsi region, Ethiopia. *Primates*, 39, 183–198. <https://doi.org/10.1007/BF02557730>
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, 74(6), 1111–1120. <https://doi.org/10.1086/421051>
- Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappel, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., & Barillot, E. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics (Oxford, England)*, 28(3), 423–425. <https://doi.org/10.1093/bioinformatics/btr670>
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Bouckaert, R., & Xie, D. (2017). BEAST2-Dev/substmodels: Standard Nucleotide Substitution Models v1.0.1 (v1.0.1). Zenodo. <https://doi.org/10.5281/zenodo.995740>
- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2011). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21(4), 481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>
- Bruvo, R., Michiels, N. K., D'Souza, T. G., & Schulenburg, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, 13(7), 2101–2106. <https://doi.org/10.1111/j.1365-294X.2004.02209.x>
- Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y., Žegarac, A., Hofmanová, Z., Winkelbach, L., Reyna-Blanco, C. S., Bieker, V., Orschiedt, J., Brinker, U., Scheu, A., Leuenberger, C., Bertino, T. S., Bollongino, R., Lidke, G., ... Wegmann, D. (2020). Low prevalence of lactase persistence in bronze age Europe indicates ongoing strong selection over the last 3,000 years. *Current Biology: CB*, 30(21), 4307–4315.e13. <https://doi.org/10.1016/j.cub.2020.08.033>
- Cahill, J. A., Soares, A. E., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1699), 20150138. <https://doi.org/10.1098/rstb.2015.0138>
- Carotenuto, F., Di Febraro, M., Melchionna, M., Castiglione, S., Saggese, F., Serio, C., Mondanaro, A., Passaro, F., Loy, A., & Raia, P. (2016). The influence of climate on species distribution over time and space during the late quaternary. *Quaternary Science Reviews*, 149, 188–199. <https://doi.org/10.1016/j.quascirev.2016.07.036>
- Casas-Gallego, M., Hahn, K., Neumann, K., Demissew, S., Schmidt, M., Bodin, S. C., & Bruch, A. A. (2023). Publisher correction: Cooling-induced expansions of Afromontane forests in the horn of Africa since the last glacial maximum. *Scientific Reports*, 13(1), 10687. <https://doi.org/10.1038/s41598-023-37919-y>
- Chala, D., Roos, C., Svenning, J. C., & Zinner, D. (2019). Species-specific effects of climate change on the distribution of suitable baboon habitats – ecological niche modeling of current and last glacial maximum conditions. *Journal of Human Evolution*, 132, 215–226. <https://doi.org/10.1016/j.jhevol.2019.05.003>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chiou, K. L., Janiak, M. C., Schneider-Crease, I. A., Sen, S., Ayele, F., Chuma, I. S., Knauf, S., Lemma, A., Signore, A. V., D'Ippolito, A. M., Abebe, B., Haile, A. A., Kebede, F., Fashing, P. J., Nguyen, N., McCann, C., Houck, M. L., Wall, J. D., Burrell, A. S., ... Snyder-Mackler, N. (2022). Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. *Nature Ecology & Evolution*, 6(5), 630–643. <https://doi.org/10.1038/s41559-022-01703-4>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- de Beaux, O. (1925). Mammiferi dell'Abissinia raccolti dal Signor Ugo Ignesti. *Atti della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale in Milano*, 64, 196–218.
- De Jong, Y. A., Moges, A., Jenner, T., Abu, K., Jenner, T., Atickem, A., Zinner, D., & Butynski, T. M. (2023). *Theropithecus gelada* ssp. Arsi. The IUCN Red List of Threatened Species 2023 [provisional].
- Delson, E. (1993). The fossils from Africa and India and the taxonomy of the genus. In N. G. Jablonski (Ed.), *Theropithecus: Rise and fall of a primate genus* (pp. 157–190). Cambridge University Press.
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., d'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D. V., Dubuis, A., Salamin, N., & Guisan, A. (2017). Ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40(6), 774–787.
- Dolotovskaya, S., Torroba Bernaldo, J., Haus, T., Noll, A., Hofreiter, M., Zinner, D., & Roos, C. (2017). Comparing mitogenomic timetrees for two African savannah primate genera (*Chlorocebus* and *Papio*). *Zoological Journal of the Linnean Society*, 181, 471–483.
- Dunbar, R. I. M. (1977). Feeding ecology of gelada baboons: A preliminary report. In T. H. Clutton-Brock (Ed.), *Primate ecology: Studies of feeding and ranging behaviour in lemurs, monkeys and apes* (pp. 251–273). Academic Press.
- Dunbar, R. I. M. (1980). Demographic and life history variables of a population of gelada baboons (*Theropithecus gelada*). *Journal of Animal Ecology*, 49(2), 485–506.
- Dunbar, R. I. M. (1993). Socioecology of the extinct theropithecids: A modelling approach. In N. G. Jablonski (Ed.), *Theropithecus: Rise and fall of a primate genus* (pp. 157–190). Cambridge University Press.
- Dunbar, R. I. M., & Dunbar, P. (1974). On hybridization between *Theropithecus gelada* and *Papio anubis* in the wild. *Journal of Human Evolution*, 3, 3–192. [https://doi.org/10.1016/0047-2484\(74\)90176-6](https://doi.org/10.1016/0047-2484(74)90176-6)
- Evershed, R. P., Davey Smith, G., Roffet-Salque, M., Timpson, A., Diekmann, Y., Lyon, M. S., Cramp, L. J. E., Casanova, E., Smyth,

- J., Whelton, H. L., Dunne, J., Brychova, V., Šoberl, L., Gerbault, P., Gillis, R. E., Heyd, V., Johnson, E., Kendall, I., Manning, K., ... Thomas, M. G. (2022). Dairying, diseases and the evolution of lactase persistence in Europe. *Nature*, 608(7922), 336–345. <https://doi.org/10.1038/s41586-022-05010-7>
- Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpour, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szegedy, C. A., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., ... Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, 13(2), 397–406. <https://doi.org/10.1074/mcp.M113.035600>
- Fashing, P. J., Nguyen, N., Venkataraman, V. V., & Kerby, J. T. (2014). Gelada feeding ecology in an intact ecosystem at Guassa, Ethiopia: Variability over time and implications for theropithecine and hominid dietary evolution. *American Journal of Physical Anthropology*, 155(1), 1–16. <https://doi.org/10.1002/ajpa.22559>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 12–4315.
- Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., ... Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. *eLife*, 8, e50989. <https://doi.org/10.7554/eLife.50989>
- Fischer, M. L., Bachofer, F., Yost, C. L., Bludau, I. J. E., Schepers, C., Foerster, V., Lamb, H., Schäbitz, F., Asrat, A., Trauth, M. H., & Junginger, A. (2021). A Phytolith supported biosphere-hydrosphere predictive model for Southern Ethiopia: Insights into paleoenvironmental changes and human landscape preferences since the last glacial maximum. *Geosciences*, 11, 418. <https://doi.org/10.3390/geosciences11100418>
- Foerster, V., Junginger, A., Langkamp, O., Gebru, T., Asrat, A., Umer, M., Lamb, H. F., Wennrich, V., Rethemeyer, J., Nowaczyk, N., Trauth, M. H., & Schaebitz, F. (2012). Climatic change recorded in the sediments of the Chew Bahir basin, Southern Ethiopia, during the last 45,000 years. *Quaternary International*, 274, 25–37. <https://doi.org/10.1016/j.quaint.2012.06.028>
- Getahun, D. A., Delson, E., & Seyoum, C. M. (2023). A review of *Theropithecus oswaldi* with the proposal of a new subspecies. *Journal of Human Evolution*, 180, 103373. <https://doi.org/10.1016/j.jhevol.2023.103373>
- Gippoliti, S. (2010). *Theropithecus gelada* distribution and variations related to taxonomy: History, challenges and implications for conservation. *Primates: Journal of Primatology*, 51(4), 291–297. <https://doi.org/10.1007/s10329-010-0202-x>
- Gippoliti, S. (2020). Everything mammal conservation biologists always wanted to know about taxonomy (but were afraid to ask). *Journal for Nature Conservation*, 54, 125793. <https://doi.org/10.1016/j.jnc.2020.125793>
- Gippoliti, S. (2022). Taxonomic history affects current conservation efforts: Examples from the primates of the horn of Africa. *Journal of Animal Diversity*, 4(1), 18–20. <https://doi.org/10.52547/JAD.2022.4.1.2>
- Gippoliti, S., Mekonnen, A., Burke, R., Nguyen, N., & Fashing, P. J. (2019). *Theropithecus gelada*. *The IUCN Red List of Threatened Species*.
- Girmay, T., & Dati, D. (2020). Population size, group and age structure of geladas (*Theropithecus gelada*) in escarpments of eastern Tigray, Ethiopia: Implication for conservation. *Journal of Ecology and Environment*, 44, 20. <https://doi.org/10.1186/s41610-020-00163-w>
- Girmay, T., & Tesfay, H. (2020). New geographical range of geladas (*Theropithecus gelada*) in Tigray region, Northern Ethiopia. *African Journal of Ecology*, 59(1), 277–280. <https://doi.org/10.1111/aje.12818>
- Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H. J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., ... Schook, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491(7424), 393–398. <https://doi.org/10.1038/nature11622>
- Groth, J. S., & Coop, G. (2024). The temporal and genomic scale of selection following hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 121(12), e2309168121. <https://doi.org/10.1073/pnas.2309168121>
- Groves, C. P., Cotterill, F. P. D., Gippoliti, S., Robovský, J., Roos, C., Taylor, P. J., & Zinner, D. (2017). Species definitions and conservation: A review and case studies from African mammals. *Conservation Genetics*, 18, 1247–1256. <https://doi.org/10.1007/s10592-017-0976-0>
- Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3), 276–292. <https://doi.org/10.1111/geb.12268>
- Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics*, 15(7), e1008302. <https://doi.org/10.1371/journal.pgen.1008302>
- Haileselassie, T. H., Mulualem, G., & Welegerima, K. (2023). The first large-scale record of gelada's spatial distribution and population size in Mountain chains of Tigray, Northern Ethiopia. *Heliyon*, 9(9), e19346. <https://doi.org/10.1016/j.heliyon.2023.e19346>
- Hallast, P., Maisano Delser, P., Batini, C., Zadik, D., Rocchi, M., Schempp, W., Tyler-Smith, C., & Jobling, M. A. (2016). Great ape Y chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. *Genome Research*, 26(4), 427–439. <https://doi.org/10.1101/gr.198754.115>
- Han, L., Wei, X., Liu, C., Volpe, G., Zhuang, Z., Zou, X., Wang, Z., Pan, T., Yuan, Y., Zhang, X., Fan, P., Guo, P., Lai, Y., Lei, Y., Liu, X., Yu, F., Shangguan, S., Lai, G., Deng, Q., ... Liu, L. (2022). Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. *Nature*, 604(7907), 723–731. <https://doi.org/10.1038/s41586-022-04587-3>
- Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109. <https://doi.org/10.1038/s41597-020-0453-3>
- Heuglin, T. v. (1863). Beiträge zur Zoologie Afrikas. Über einige Säugethiere des Bäschlo-Gebietes. *Nova Acta Academiae Caesarea Leopoldinae Naturae*, 30, 1–14.
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2023). Dismo: Species distribution Modeling. R Package, Version 1.3-14. <https://RAN.R-project.org/package=dismo>
- Iwamoto, T. (1993a). The ecology of *Theropithecus gelada*. In N. G. Jablonski (Ed.), *Theropithecus: Rise and fall of a primate genus* (pp. 157–190). Cambridge University Press.
- Iwamoto, T. (1993b). Food digestion and energetic conditions in *Theropithecus gelada*. In N. G. Jablonski (Ed.), *Theropithecus: Rise and fall of a primate genus* (pp. 157–190). Cambridge University Press.
- Jarvey, J. C., Low, B. S., Pappano, D. J., Bergman, T. J., & Beehner, J. C. (2018). Graminivory and fallback foods: Annual diet profile of geladas (*Theropithecus gelada*) living in the Simien Mountains national Park, Ethiopia. *International Journal of Primatology*, 39, 1–126.
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, 21(1), 241. <https://doi.org/10.1186/s13059-020-02154-5>

- Jolly, C. J., Woolley-Barker, T., Beyene, S., Disotell, T. R., & Phillips-Conroy, J. E. (1997). Intergeneric hybrid baboons. *International Journal of Primatology*, 18, 597–627. <https://doi.org/10.1023/A:1026367307470>
- Joslin, S. E. K., Durbin-Johnson, B. P., Britton, M., Settles, M. L., Korf, I., & Lemay, D. G. (2020). Association of the lactase persistence haplotype block with disease risk in populations of European descent. *Frontiers in Genetics*, 11, 558762. <https://doi.org/10.3389/fgene.2020.558762>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122. <https://doi.org/10.1038/sdata.2017.122>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Keppel, G., Van Niel, K. P., Wardell-Johnson, G. W., Yates, C. J., Byrne, M., Mucina, L., Schut, A. G. T., Hopper, S. D., & Franklin, S. E. (2012). Refugia: Identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, 21(4), 393–404.
- Kifle, Z., & Bekele, A. (2021). Feeding ecology and diet of the southern geladas (*Theropithecus gelada obscurus*) in human-modified landscape, Wollo, Ethiopia. *Ecology and Evolution*, 11(16), 11373–11386. <https://doi.org/10.1002/ece3.7927>
- King, M. (1995). Species evolution: The role of chromosome change. *Systematic Biology*, 44(4), 578–580. <https://doi.org/10.1093/sysbio/44.4.578>
- Kostin, D., Kasso, M., Komarova, V., Martynov, A., Gromov, A., Alexandrov, D., Bekele, A., Zewdie, C., Bryja, J., & Lavrenchenko, L. (2019). Taxonomic and genetic diversity of rodents from the Arsi Mountains (Ethiopia). *Mammalia*, 83(3), 237–247. <https://doi.org/10.1515/mammalia-2017-0135>
- Lavrenchenko, L. A., & Bekele, A. (2017). Diversity and conservation of Ethiopian mammals: What have we learned in 30 years? *Ethiopian Journal of Biological Sciences*, 16, 1.
- le Roux, A., Beehner, J. C., & Bergman, T. J. (2011). Female philopatry and dominance patterns in wild geladas. *American Journal of Primatology*, 73(5), 422–430. <https://doi.org/10.1002/ajp.20916>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liedigk, R., Roos, C., Brameier, M., & Zinner, D. (2014). Mitogenomics of the Old World monkey tribe Papionini. *BMC Evolutionary Biology*, 14, 176. <https://doi.org/10.1186/s12862-014-0176-1>
- Maier, R., Flegontov, P., Flegontova, O., Işildak, U., Changmai, P., & Reich, D. (2023). On the limits of fitting complex models of population history to *f*-statistics. *eLife*, 12, e85492. <https://doi.org/10.7554/eLife.85492>
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)*, 26(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
- Markarjan, D. S., Isakov, E. P., & Kondakov, G. I. (1974). Intergeneric hybrids of the lower (42-chromosome) monkey species of the Sukhumi monkey colony. *Journal of Human Evolution*, 3, 3–255. [https://doi.org/10.1016/0047-2484\(74\)90183-3](https://doi.org/10.1016/0047-2484(74)90183-3)
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2008). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15(1), 59–69. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- Mau, M., Südekum, K. H., Johann, A., Sliwa, A., & Kaiser, T. M. (2010). Indication of higher salivary alpha-amylase expression in hamadryas baboons and geladas compared to chimpanzees and humans. *Journal of Medical Primatology*, 39(3), 187–190. <https://doi.org/10.1111/j.1600-0684.2010.00407.x>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). Global biodiversity conservation: The critical role of hotspots. In F. Zachos & J. Habel (Eds.), *Biodiversity hotspots*. Springer. https://doi.org/10.1007/978-3-642-20992-5_1
- Montinaro, F., & Capelli, C. (2020). A worldwide map of human structural variants. *Trend in Genetics*, 36(10), 722–725. <https://doi.org/10.1016/j.tig.2020.07.002>
- Mori, A., & Belay, G. (1990). The distribution of baboon species and a new population of gelada baboons along the Wabi-Shebeli river, Ethiopia. *Primates*, 31, 495–508.
- Mutti, G., Oteo-Garcia, G., Caldon, M., da Silva, M. J. F., Minhós, T., Cowlshaw, G., Gottelli, D., Huchard, E., Carter, A., Martinez, F. I., Raveane, A., & Capelli, C. (2023). Assessing the recovery of Y chromosome microsatellites with population genomic data using Papio and Theropithecus genomes. *Scientific Reports*, 13(1), 13839. <https://doi.org/10.1038/s41598-023-40931-x>
- Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37, 2–203.
- Nogués-Bravo, D., Ohlemüller, R., Batra, P., & Araújo, M. B. (2010). Climate predictors of late quaternary extinctions. *Evolution: International Journal of Organic Evolution*, 64(8), 2442–2449. <https://doi.org/10.1111/j.1558-5646.2010.01009.x>
- Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R. A., Garneau, D., Daugherty, E., Globig, A., Ruhl, S., & Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *eLife*, 8, e44628. <https://doi.org/10.7554/eLife.44628>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>

- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., Kessing, B., Pontius, J., Roelke, M., Rumpel, Y., Schneider, M. P., Silva, A., O'Brien, S. J., & Pecon-Slatery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, 7(3), e1001342. <https://doi.org/10.1371/journal.pgen.1001342>
- Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, 39(10), 1256–1260. <https://doi.org/10.1038/ng2123>
- Pigkford, M. (1993). Climate change, biogeography and *Theropithecus*. In N. G. Jablonski (Ed.), *Theropithecus: The rise and fall of a primate genus* (pp. 227–244). Cambridge University Press.
- Pozzi, L., & Penna, A. (2022). Rocks and clocks revised: New promises and challenges in dating the primate tree of life. *Evolutionary Anthropology*, 31(3), 138–153. <https://doi.org/10.1002/evan.21940>
- Raia, P., Passaro, F., Fulgione, D., & Carotenuto, F. (2012). Habitat tracking, stasis and survival in Neogene large mammals. *Biology Letters*, 8(1), 64–66. <https://doi.org/10.1098/rsbl.2011.0613>
- Roos, C., Knauf, S., Chuma, I. S., Maille, A., Callou, C., Sabin, R., Portela Miguez, R., & Zinner, D. (2021). New mitogenomic lineages in Papio baboons and their phylogeographic implications. *American Journal of Physical Anthropology*, 174(3), 407–417. <https://doi.org/10.1002/ajpa.24186>
- Rüppell, W. P. E. S. (1835). *Neue Wirbelthiere zu der Fauna von Abyssinien gehörig, Säugethiere*. Siegmund Schmerber.
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909), 832–837. <https://doi.org/10.1038/nature01140>
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., ... Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. <https://doi.org/10.1038/nature06250>
- Samuelson, L. C., Phillips, R. S., & Swanberg, L. J. (1996). Amylase gene structures in primates: Retroposon insertions and promoter evolution. *Molecular Biology and Evolution*, 13(6), 767–779. <https://doi.org/10.1093/oxfordjournals.molbev.a025637>
- Santander, C., Molinaro, L., Mutti, G., Martínez, F. I., Mathe, J., Ferreira da Silva, M. J., Caldon, M., Oteo-García, G., Aldeias, V., Archer, W., Bamford, M., Biro, D., Bobe, R., Braun, D. R., Hammond, P., Lüdecke, T., Pinto, M. J., Meira Paulo, L., Stalmans, M., ... Capelli, C. (2022). Genomic variation in baboons from central Mozambique unveils complex evolutionary relationships with other Papio species. *BMC Ecology and Evolution*, 22(1), 44. <https://doi.org/10.1186/s12862-022-01999-7>
- Schoener, T. W. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, 51, 3–418.
- Shotake, T., Saijuntha, W., Agatsuma, T., & Kawamoto, Y. (2016). Genetic diversity within and among gelada (*Theropithecus gelada*) populations based on mitochondrial DNA analysis. *Anthropological Science*, 124, 157–167. <https://doi.org/10.1537/ase.160717>
- Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., Storer, J. M., Kuhlwilm, M., Fontseré, C., Seshadri, L., Bergey, C. M., Burrell, A. S., Bergman, J., Phillips-Conroy, J. E., Shiferaw, F., Chiou, K. L., Chuma, I. S., Keyyu, J. D., Fischer, J., ... Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science (New York, N.Y.)*, 380(6648), eabn8153. <https://doi.org/10.1126/science.abn8153>
- Stewart, B. M., Turner, S. E., & Matthews, H. D. (2020). Climate change impacts on potential future ranges of non-human primate species. *Climatic Change*, 162, 2301–2318.
- Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B. P., Baker, C., Nordenfelt, S., Bamshad, M., Jorde, L. B., Posukh, O. L., Sahakyan, H., Watkins, W. S., Yepiskoposyan, L., Abdullah, M. S., Bravi, C. M., Capelli, C., Hervig, T., ... Eichler, E. E. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science (New York, N.Y.)*, 349(6253), aab3761. <https://doi.org/10.1126/science.aab3761>
- Svenning, J. C., Fløjgaard, C., Marske, K. A., Nogueira-Bravo, D., & Normand, S. (2011). Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, 30(21–22), 2930–2947.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science (New York, N.Y.)*, 240(4857), 1285–1293. <https://doi.org/10.1126/science.3287615>
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, 32(3), 369–373. <http://www.jstor.org/stable/30243260>
- Timmermann, A., & Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. *Nature*, 538(7623), 92–95. <https://doi.org/10.1038/nature19365>
- Timmermann, A., Yun, K. S., Raia, P., Ruan, J., Mondanaro, A., Zeller, E., Zollikofer, C., Ponce de León, M., Lemmon, D., Willeit, M., & Ganopolski, A. (2022). Climate effects on archaic human habitats and species successions. *Nature*, 604(7906), 495–501. <https://doi.org/10.1038/s41586-022-04600-9>
- Tobler, R., Souilmi, Y., Huber, C. D., Bean, N., Turney, C. S. M., Grey, S. T., & Cooper, A. (2023). The role of genetic selection and climatic factors in the dispersal of anatomically modern humans out of Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 120(22), e2213061120. <https://doi.org/10.1073/pnas.2213061120>
- Torres, S. M. F., Furrow, E., Souza, C. P., Granick, J. L., de Jong, E. P., Griffin, T. J., & Wang, X. (2018). Salivary proteomics of healthy dogs: An in depth catalog. *PLoS One*, 13(1), e0191307. <https://doi.org/10.1371/journal.pone.0191307>
- Trauth, M. H., Deino, A. L., Bergner, A. G., & Strecker, M. R. (2003). East African climate change and orbital forcing during the last 175 kyr BP. *Earth and Planetary Science Letters*, 206(3–4), 297–313.
- Trede, F., Lemkul, A., Atickem, A., Beehner, J. C., Bergman, T. J., Burke, R. J., Fashing, P. J., Knauf, S., Mekonnen, A., Moges, A., Nguyen, N., Roos, C., & Zinner, D. (2020). Geographic distribution of microsatellite alleles in geladas (primates, Cercopithecidae): Evidence for three evolutionary units. *Zoologica Scripta*, 49, 659–667.
- Vasimuddin, M., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil (pp. 314–324).
- Vivero, J. L., Ensermu, K., & Sebsebe, D. (2006). Progress on the red list of plants of Ethiopia and Eritrea: Conservation and biogeography of endemic flowering taxa. In S. A. Ghazanfar & H. J. Beentje (Eds.), *Taxonomy and ecology of African plants, their conservation and sustainable use* (pp. 761–778). Royal Botanic Gardens, Kew.
- Walker, J. A., Jordan, V. E., Storer, J. M., Steely, C. J., Gonzalez-Quiroga, P., Beckstrom, T. O., Rewerts, L. C., St Romain, C. P., Rockwell, C. E., Rogers, J., Jolly, C. J., Konkel, M. K., Baboon Genome Analysis Consortium, & Batzer, M. A. (2019). *Alu* insertion polymorphisms shared by *Papio* baboons and *Theropithecus gelada* reveal an intertwined common ancestry. *Mobile DNA*, 10, 46. <https://doi.org/10.1186/s13100-019-0187-y>
- Warren, W. C., Harris, R. A., Haukness, M., Fiddes, I. T., Murali, S. C., Fernandes, J., Dishuck, P. C., Storer, J. M., Raveendran, M., Hillier, L. W., Porubsky, D., Mao, Y., Gordon, D., Vollger, M. R., Lewis, A. P., Munson, K. M., DeVogelaere, E., Armstrong, J., Diekhans, M., ... Eichler, E. E. (2020). Sequence diversity analyses of an improved

- rhesus macaque genome enhance its biomedical utility. *Science* (New York, N.Y.), 370(6523), eabc6617. <https://doi.org/10.1126/science.abc6617>
- Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, 399(6737), 682–685. <https://doi.org/10.1038/21415>
- Wilson Sayres, M. A., Venditti, C., Pagel, M., & Makova, K. D. (2011). Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution: International Journal of Organic Evolution*, 65(10), 2800–2815. <https://doi.org/10.1111/j.1558-5646.2011.01337.x>
- Woldegeorgis, C., & Bekele, A. (2015). Activity budget and behavioural patterns of gelada *Theropithecus gelada* (Mammalia: Primates: Cercopithecidae) on the Gich plateau of the Simien Mountains National Park, Ethiopia. *Journal of Threatened Taxa*, 7(8), 7415.
- Wu, F. L., Strand, A. I., Cox, L. A., Ober, C., Wall, J. D., Moorjani, P., & Przeworski, M. (2020). A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biology*, 18(8), e3000838. <https://doi.org/10.1371/journal.pbio.3000838>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, 2(3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Yalden, D. W., Lagen, M. J., & Kock, D. (1976). Catalogue of the mammals of Ethiopia. *Monitore Zoologico Italiano*, 8, 1–118.
- Yalden, D. W., Lagen, M. J., Kock, D., & Hillman, J. C. (1996). Catalogue of the mammals of Ethiopia and Eritrea. 7. Revised checklist, zoogeography and conservation. *Tropical Zoology*, 9(1), 73–164. <https://doi.org/10.1080/03946975.1996.10539304>
- Zinner, D., Atickem, A., Beehner, J. C., Bekele, A., Bergman, T. J., Burke, R., Dolotovskaya, S., Fashing, P. J., Gippoliti, S., Knauf, S., Knauf, Y., Mekonnen, A., Moges, A., Nguyen, N., Stenseth, N. C., & Roos, C. (2018). Phylogeography, mitochondrial DNA diversity, and demographic history of geladas (*Theropithecus gelada*). *PLoS One*, 13(8), e0202303. <https://doi.org/10.1371/journal.pone.0202303>
- Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American Journal of Physical Anthropology*, 150(1), 133–140. <https://doi.org/10.1002/ajpa.22185>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 1–14.

SUPPORTING INFORMATION

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

How to cite this article: Caldon, M., Mutti, G., Mondanaro, A., Imai, H., Shotake, T., Oteo Garcia, G., Belay, G., Morata, J., Trotta, J.-R., Montinaro, F., Gippoliti, S., & Capelli, C. (2024). Gelada genomes highlight events of gene flow, hybridisation and local adaptation that track past climatic changes. *Molecular Ecology*, 33, e17514. <https://doi.org/10.1111/mec.17514>

Supplemental Information for:

Gelada genomes highlight events of gene flow, hybridisation and local adaptation that track past climatic changes.

Matteo Caldon, Giacomo Mutti, Alessandro Mondanaro, Hiroo Imai, Takayoshi Shotake, Gonzalo Oteo Garcia, Gurja Belay, Jordi Morata, Jean-Rémi Trotta, Francesco Montinaro, Spartaco Gippoliti, Cristian Capelli*

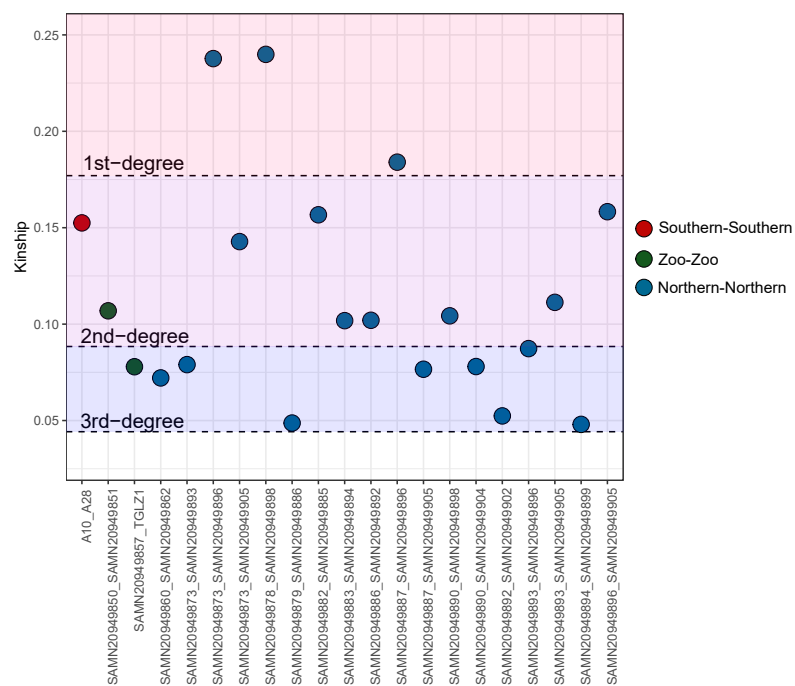
*Corresponding author. Email: cristian.capelli@unipr.it

This PDF file includes:

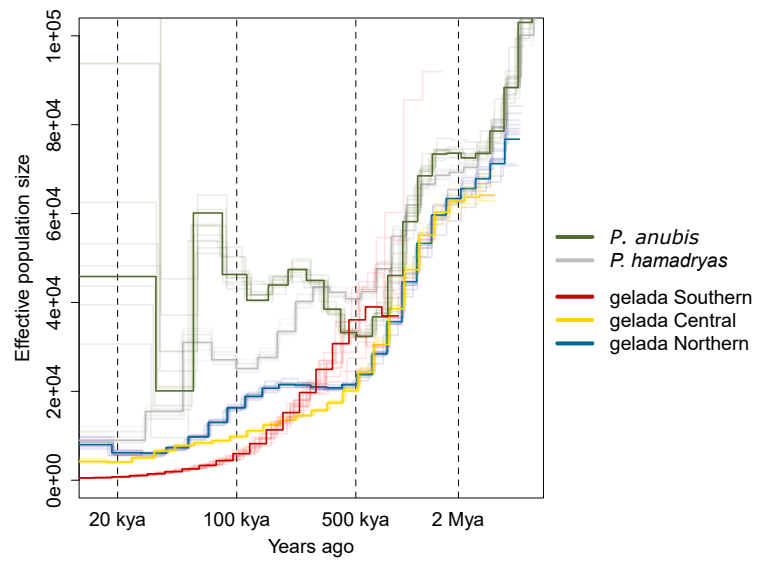
- Supplementary Figures 1 to 10

Other Supplementary Material for this manuscript includes the following:

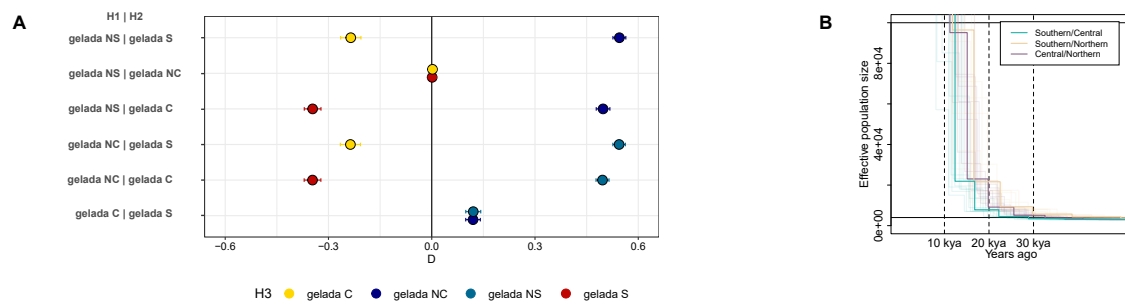
- STables 1 to 6.
- Supplementary Material 1 (GIF).



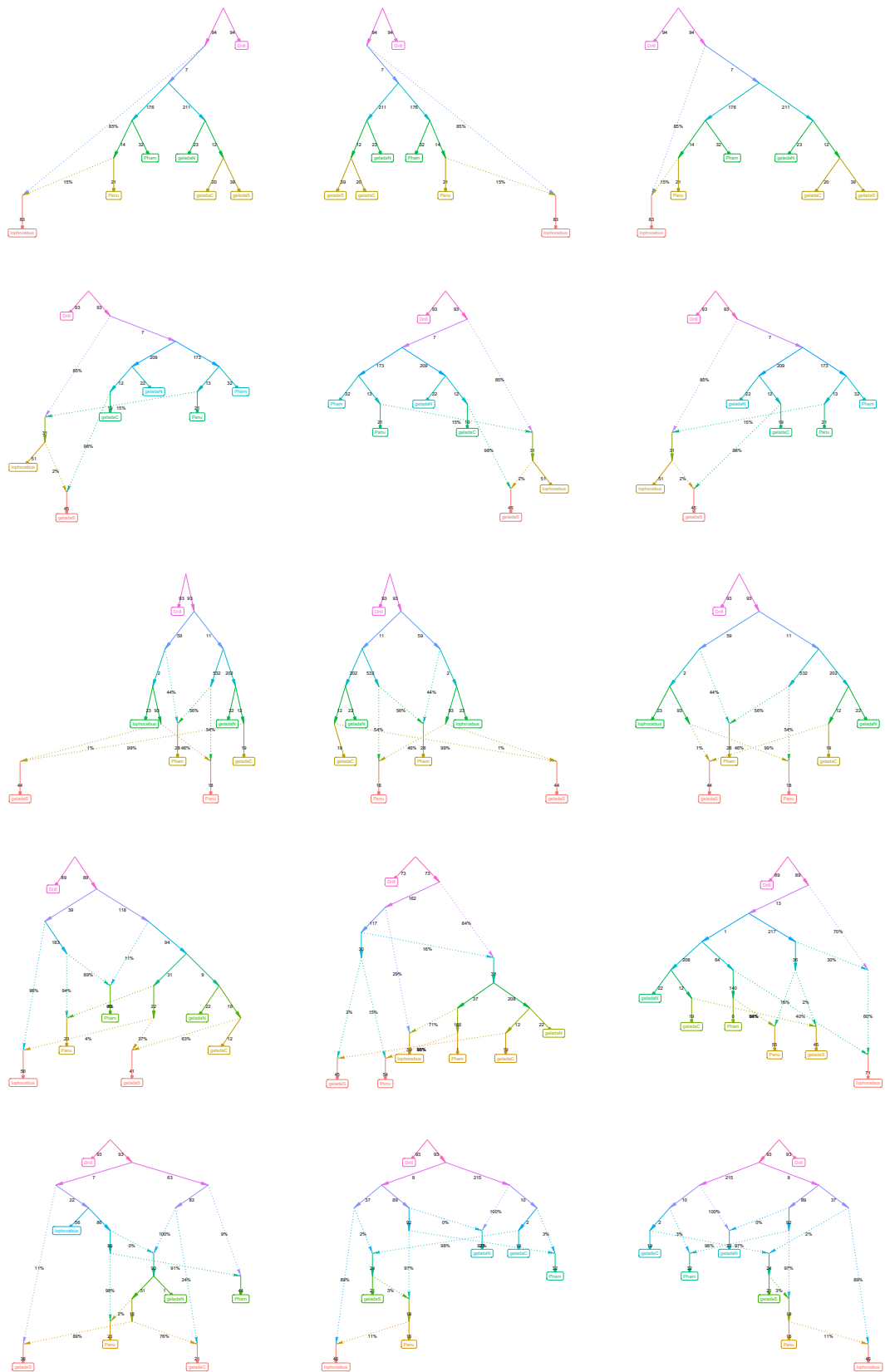
Supplementary Figure 2 - Kinship analysis. Pairs of gelada individuals with kinship relationships up to the third degree.



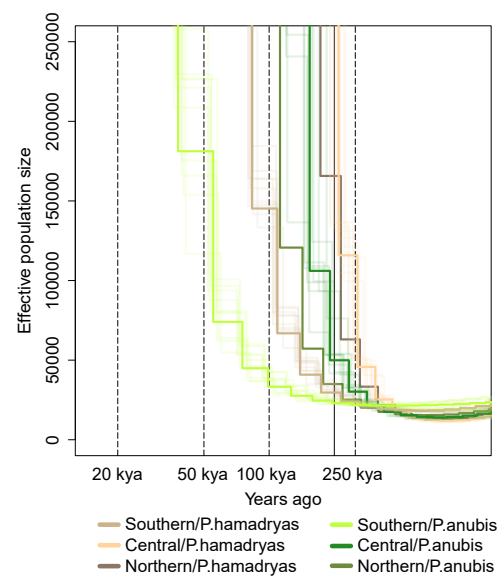
Supplementary Figure 3 - PSMC curves for *Theropithecus* and *Papio*. Changes in effective population size across time estimated using PSMC for *T. gelada* populations, *P. anubis* and *P. hamadryas*. X-axis on logarithmic scale.



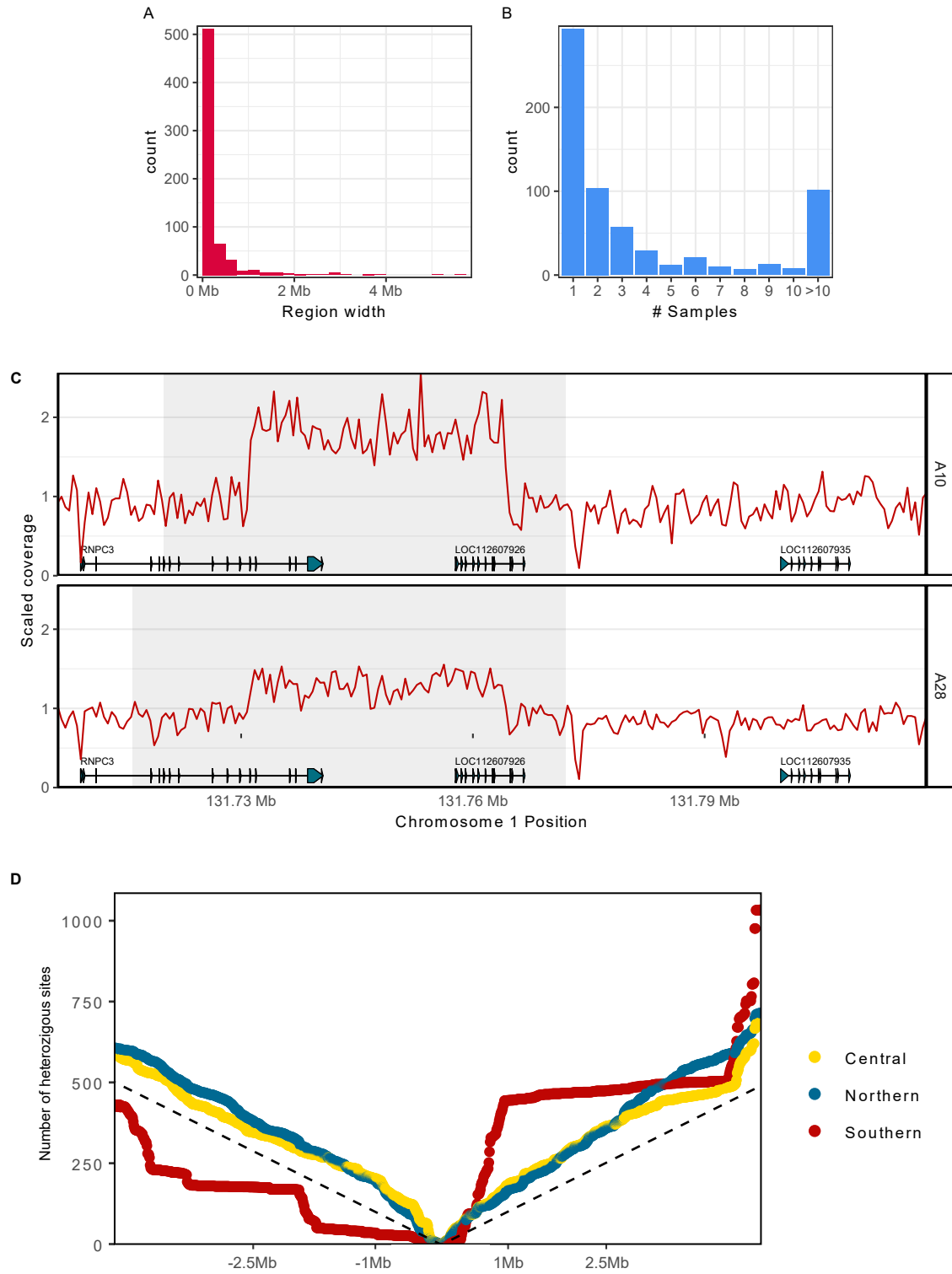
Supplementary Figure 4 - Relationship among gelada populations. A) ABBA-BABA test between gelada populations (Central as C, Southern as S, Northern split as NS - Sankaber locality - and NC - Chenek locality); B) coalescence of artificial X chromosome hybrids (hPSMC analysis) of SAMN20949878, SAMN20949865, and A28 genomes mapped on the alternative reference genome *Mmul_10*.



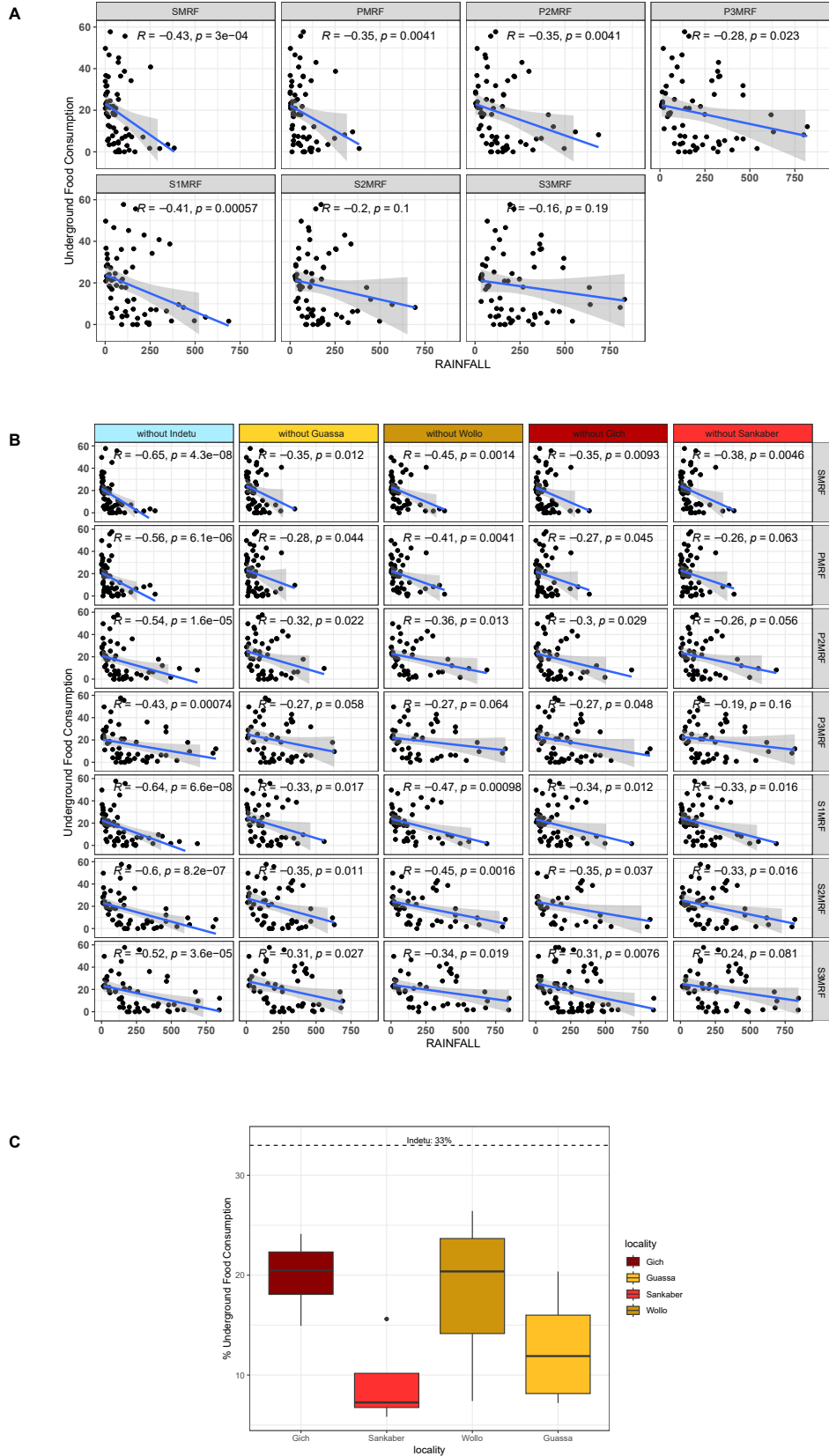
Supplementary Figure 5 - Admixture graphs: Gelada, baboons and *L. aethiops* relationships. Admixture graphs reporting relationships and admixture events between Papionines species, considering 1 to 5 admixture events (from top to bottom); for each set of considered admixture events, the three topologies with the best scores are reported (from left to right). Dotted arrows indicate admixture, while associated numbers report the amount of each contribution; continuous arrows indicate linear descendants, while associated numbers are indicative of the amount of drift experienced along lineages given in f-statistics units multiplied by 1,000.



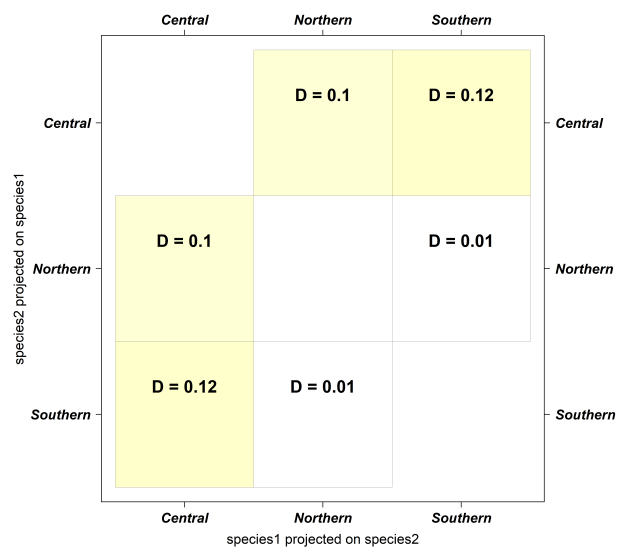
Supplementary Figure 6 - hPSMC analysis on *Papio-Theropithecus*, alternative reference genome. Coalescence of artificial X chromosome *Papio-Theropithecus* hybrids with genomes mapped on *Tgel_1.0* reference genome; the x-axis is in logarithmic scale.



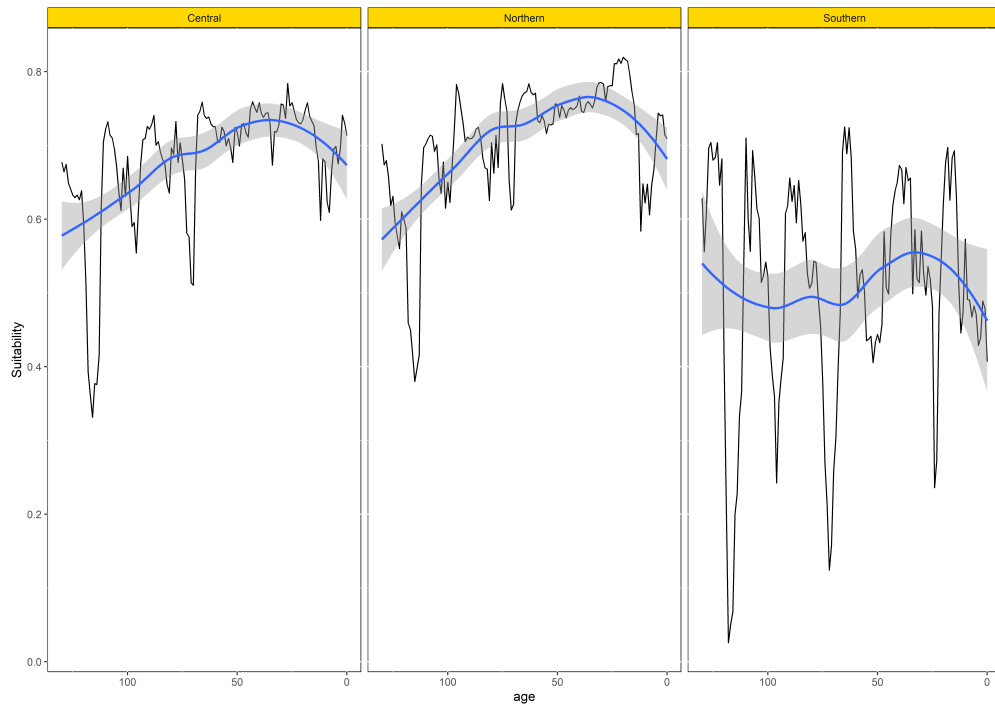
Supplementary Figure 7 - CNVs and amylase gene copy number. A) Size distribution of expanded regions; B) Frequency spectrum across the whole gelada dataset; C) amylase duplication extent; D) heterozygosity pattern estimates in 10kb-windows over a region of 10Mb centred around the two pancreas amylase loci (LOC112607935 and LOC112607926) for each gelada populations.



Supplementary Figure 8 - Underground Food consumption and rainfalls. Correlation analysis for Underground Food consumption and rainfall using A) data from all 5 localities described in the main text and B) different subsets generated by removing data from one locality at a time. Correlation analyses are ordered by columns according to the dataset removed and by rows considering different rainfall estimates. In each panel the regression coefficient (R) and the p -value (p) are reported; acronyms should be read as follows: SMRF = same month rainfalls, PMRF = previous month rainfalls, P2MRF = previous two month rainfalls, P3MRF = previous three month rainfalls, S1MRF = same month + previous month rainfalls, S2MRF = same month + previous two month rainfalls, S3MRF = same month + previous three month rainfalls. C) Box-plots reporting the distribution of UF consumption estimates calculated averaging across a moving window of seven consecutive months. The dotted line refers to the mean value across the seven continuous months available for Indetu.



Supplementary Figure 9 - Pairwise niche similarity between each gelada population quantified using the Schoener's D metric.



Supplementary Figure 10 - Past habitat suitability values averaged over the current occurrence of gelada population, for the last 130 ka. The blue lines indicate the LOESS regression curves fitted over time. The grey bands represent the 95% confidence interval around LOESS estimates.

Supplementary Table 4 - Genomic location of population-specific CNVs, their gene content and number of individuals with more than 2 copies per genome.

chr	start	end	Entrez_geneid	gene_name	n	range	population
1	22884000	22919999	NA	NA	16	2-3	Northern
1	41480000	41507999	112609905, 112609443, 112622928	PRAMEF12, LOC112609443, LOC112622928	9	2-3	Central
1	131716000	131771999	112607915, 112607926	RNPC3, LOC112607926	2	3-4	Southern
1	178200000	178275999	NA	NA	14	2-4	Northern
1	222920000	222963999	112632947	PGBD2	15	2-4	Northern
2	95904000	95939999	112619547	C2H3orf86	18	2-3	Northern
2	140436000	140527999	112618631	AADACL2	15	2-3	Northern
3	256000	287999	NA	NA	2	3-4	Southern
3	676000	959999	NA	NA	49	3-4	Northern
3	44404000	44447999	NA	NA	6	2-3	Central
3	55012000	55091999	112620654	LOC112620654	2	3-4	Southern
3	111420000	111479999	112621486, 112620677, 112621313	FAM133B, RBM48, PEX1	30	2-3	Northern
4	34896000	34943999	112622088	LOC112622088	24	2-6	Northern
4	77600000	77643999	NA	NA	14	2-4	Northern
5	30884000	30931999	NA	NA	13	2-3	Northern
5	33252000	33287999	NA	NA	2	4-4	Southern
7a	9844000	9887999	112627372, 112628047	CATSPER2, LOC112628047	5	2-3	Central
7b	488000	515999	NA	NA	8	2-3	Central
7b	9936000	9991999	NA	NA	9	2-4	Central
7b	65692000	65731999	NA	NA	5	2-4	Central
8	336000	399999	112630476	LOC112630476	2	3-3	Southern
8	42732000	42827999	NA	NA	6	2-3	Central
8	46596000	46683999	NA	NA	39	2-4	Northern
8	46784000	46855999	112630152	RNF4	46	2-3	Northern
8	46932000	46999999	NA	NA	10	2-4	Central
8	96868000	96907999	112630312	VIRMA	6	2-3	Central
9	12792000	12867999	NA	NA	17	2-3	Northern
9	34604000	34691999	112631349, 112631348	ACBD7, OLAH	16	2-5	Northern
10	33928000	33979999	NA	NA	7	2-3	Central
10	34040000	34083999	112632805	LOC112632805	14	2-4	Central
11	93404000	93611999	112634982	LOC112634982	15	2-4	Northern
11	129096000	129131999	NA	NA	7	2-4	Central
12	69716000	69751999	NA	NA	5	2-3	Central
13	49308000	49379999	112605158	WDR43	5	2-4	Central
14	62956000	62995999	112607182, 112607346	LOC112607182, LOC112607346	41	2-3	Northern
14	67316000	67363999	112605568	RPS3	2	3-3	Southern
14	120776000	120847999	NA	NA	6	2-3	Central
15	46496000	46527999	112608312, 112608310, 112608313, 112608309	RGP1, GBA2, CREB3, TLN1	15	2-3	Northern
15	111224000	111259999	NA	NA	2	4-4	Southern
16	47756000	47795999	112609529	CD300A	2	3-4	Southern
17	19524000	19567999	112611151	SPRY2	6	2-3	Central
17	26956000	26995999	NA	NA	2	3-3	Southern
18	39244000	39279999	NA	NA	2	3-4	Southern
19	9040000	9091999	112613222, 112613197	LOC112613222, LOC112613197	16	2-3	Northern
19	55488000	55535999	NA	NA	36	2-4	Northern
19	57488000	57527999	NA	NA	20	2-4	Northern
20	2344000	2407999	NA	NA	6	2-3	Central