ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

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

Matteo Caldon[1] | Giacomo Mutti[1,2,3] | Alessandro Mondanaro[4] | Hiroo Imai[5] |
Takayoshi Shotake[6] | Gonzalo Oteo Garcia[1,7,8] | Gurja Belay[9] | Jordi Morata[10] |
Jean-Rémi Trotta[10] | Francesco Montinaro[11,12] | Spartaco Gippoliti[13,14] |
Cristian Capelli[1,15]

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

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

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

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

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

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

[7]Centre for Palaeogenetics, Stockholm, Sweden

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

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

[10]Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain

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

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

[13]IUCN/SSC Primate Specialist Group, Rome, Italy

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

[15]Department 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

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.

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
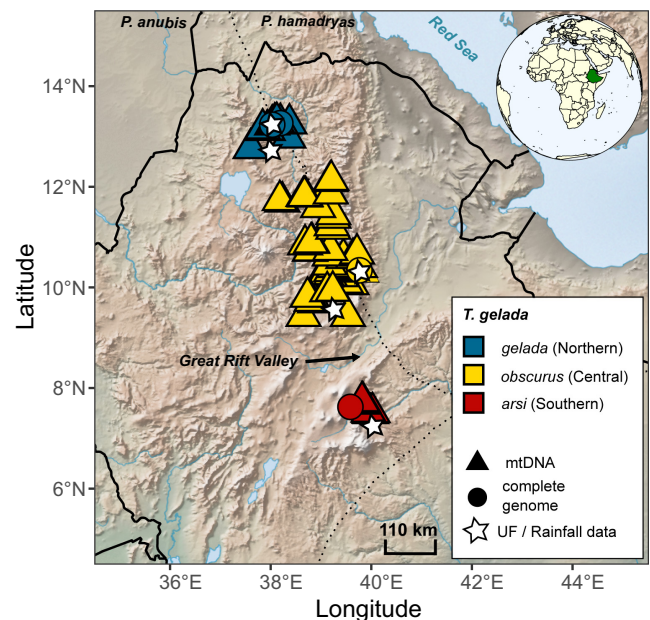
FIGURE 1 Geographical origin in the Ethiopian highlands of *Theropithecus gelada* samples analysed in this work. Colours refer to source populations (Northern, Central and Southern) and shapes identify the data available for the sample (Complete genome sequence and mitochondrial DNA only). The dotted 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.

—WILEY

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 climactic 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 30×, while sample A10 reached a median coverage of 12×, 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 33.65×. 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×. 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-Theropitecus* (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 100 bp 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±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 *smartpca* 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 *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.

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 intraspecies 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 10 kb; 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 rs 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 2 Ma CESM1.2 simulation (Timmermann et al., 2022) downscaled at 0.5°×0.5° 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 2 Ma 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 km × 1 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). Schoner' 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 TMRCAs of the mitochondrial DNA phylogeny was about twice the age of those recovered from the Y chromosome (Figure 2a,b). Substantial differences in TMRCAs between the Y chromosome and mtDNA have been reported for gorillas and orangutans, but not for chimpanzees and humans (Hallast et al., 2016).
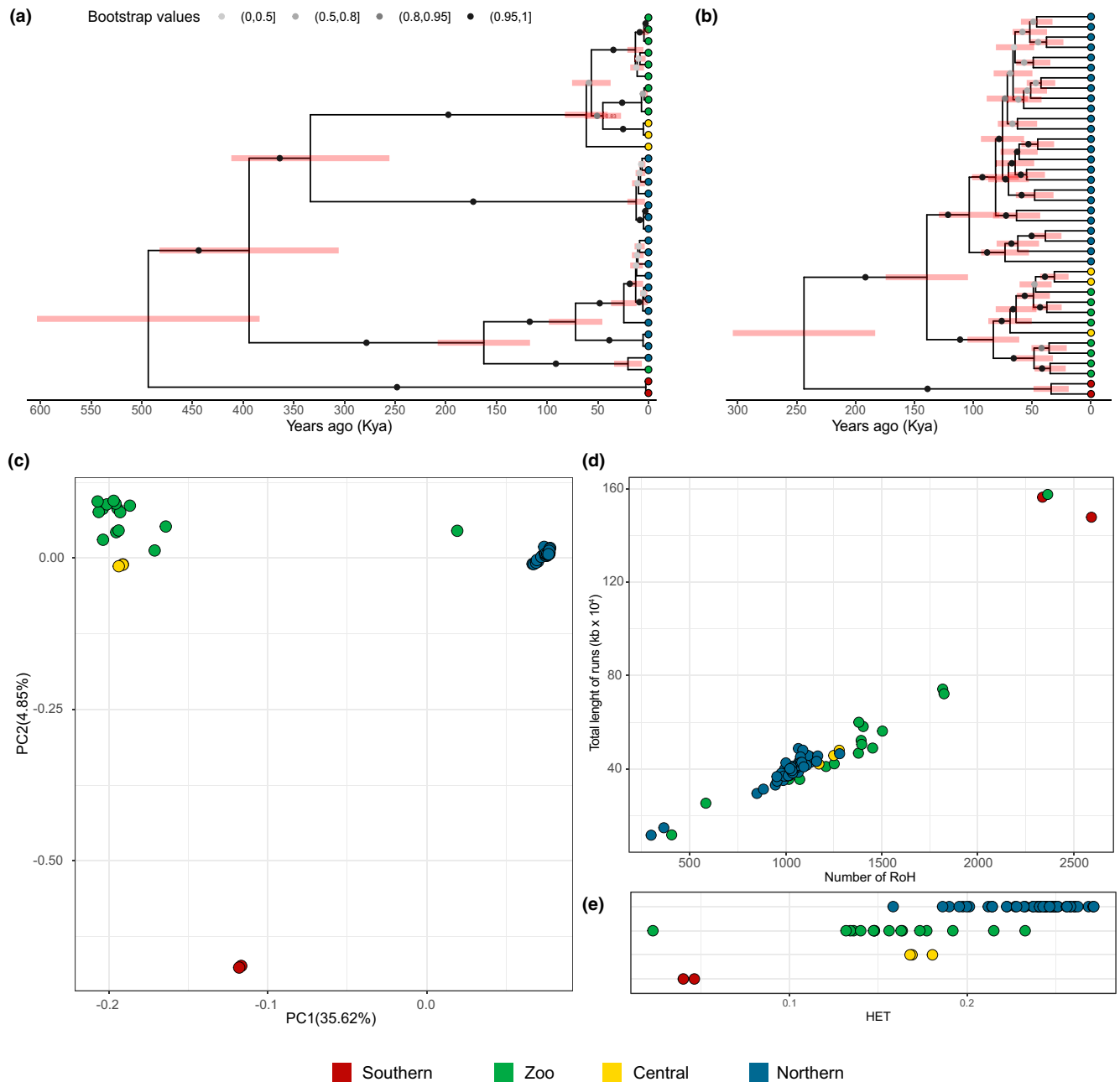
FIGURE 2 Population structure of *Theropithecus gelada*. Colours referred to gelada taxa as in Figure 1, with the addition of samples from zoos, in green. (a) Whole mitogenomes and (b) Y chromosome sequence phylogenetic trees; the red bars refer to highest posterior density (see Table S2), and the grey 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 1.29× is also much lower than the rest of the samples here analysed (mean 14.5×; 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
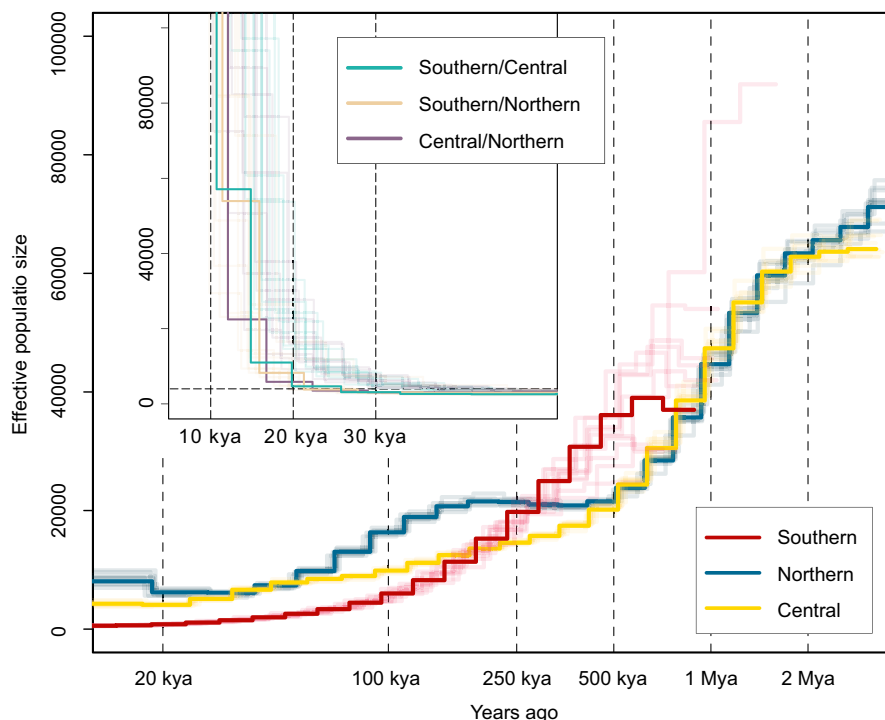

FIGURE 3 Demographic history of gelada populations. Changes in effective population size across time estimated using *PSMC*. SAMN20949878, SAMN20949865 and A28 genomes were analysed for Northern Central and Southern geladas respectively. *X*-axis on logarithmic scale. Colours as in Figure 1. Inset: hPMSC coalescence curves of artificial X chromosome hybrids of the same three samples, mapped on Tgel_1.0 reference genome.
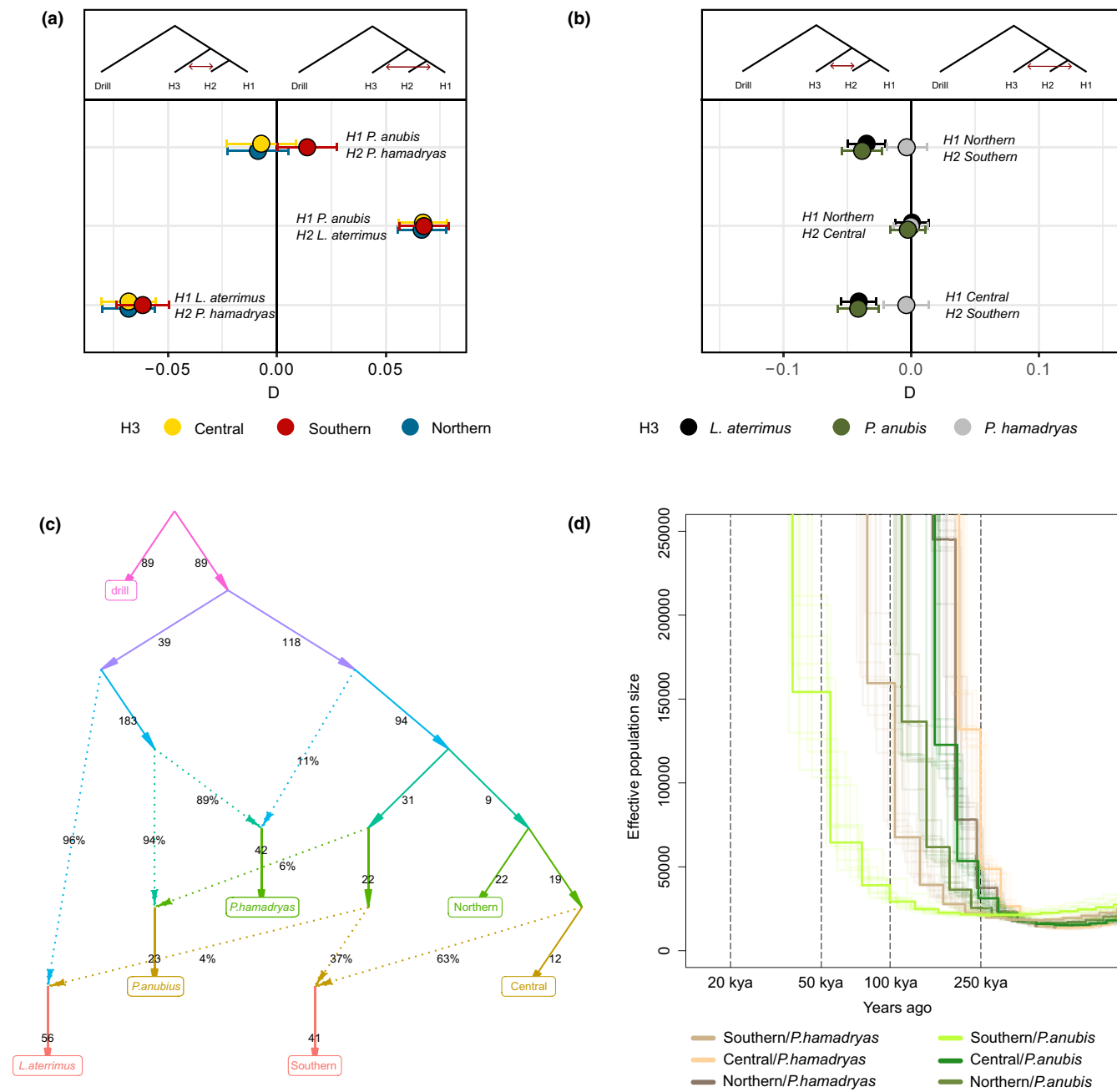
FIGURE 4 Gelada-baboons hybridisation events. (a) ABBA–BABA tests comparing *Theropithecus gelada* populations (H3) and baboons sp. (as H1 and H2); (b) ABBABABA test comparing baboons sp./*Lophocebus aterrimus* (as H3) and *T. gelada* populations (as H1 and H2); bars show the extent of three standard 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; 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. aterrimus* (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
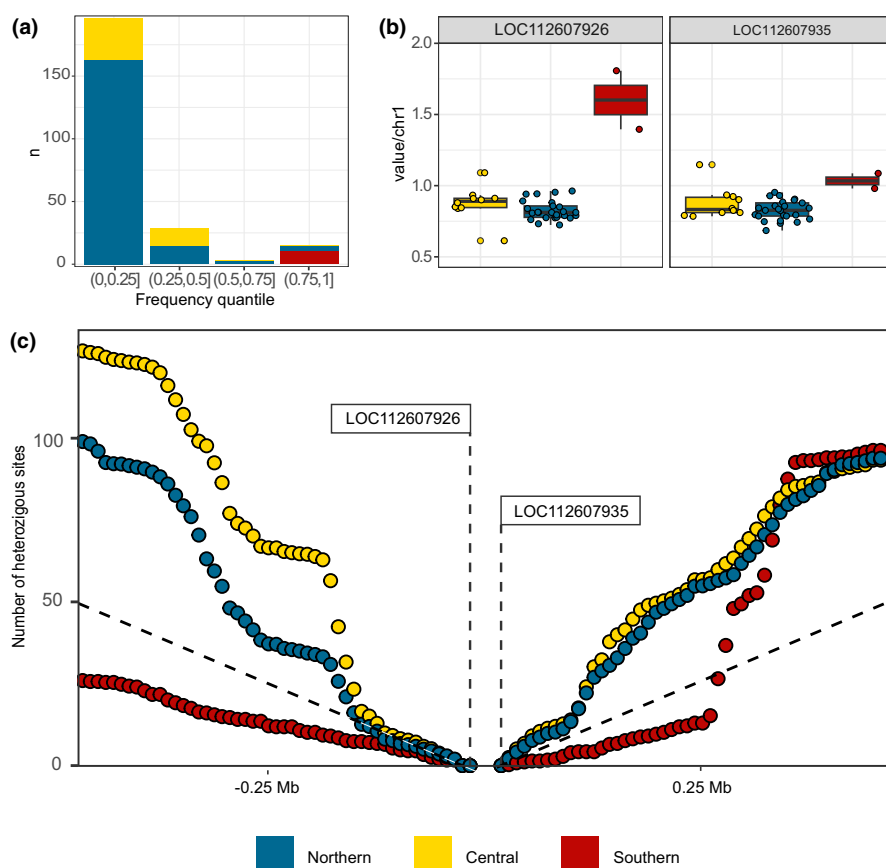

FIGURE 5 CNVs and amylase gene copy number. (a) Population-specific CNVs divided according to their frequency in each population. (b) Normalised genomic coverage over the two loci annotated as pancreatic amylase in geladas. (c) Cumulative number of heterozygous positions in 10-kb windows, starting from 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 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). 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, $rs = -.43126$, p(2-tailed) $= .0003$, monthly rainfall during the previous month, $rs = -.349$, p(2-tailed) $= .00408$, cumulative rainfall during previous 2 months, $rs = -.34862$, p(2-tailed) $= .00412$ and cumulative rainfall during previous 3 months, $rs = -.2801$, p(2-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 suitablities 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).
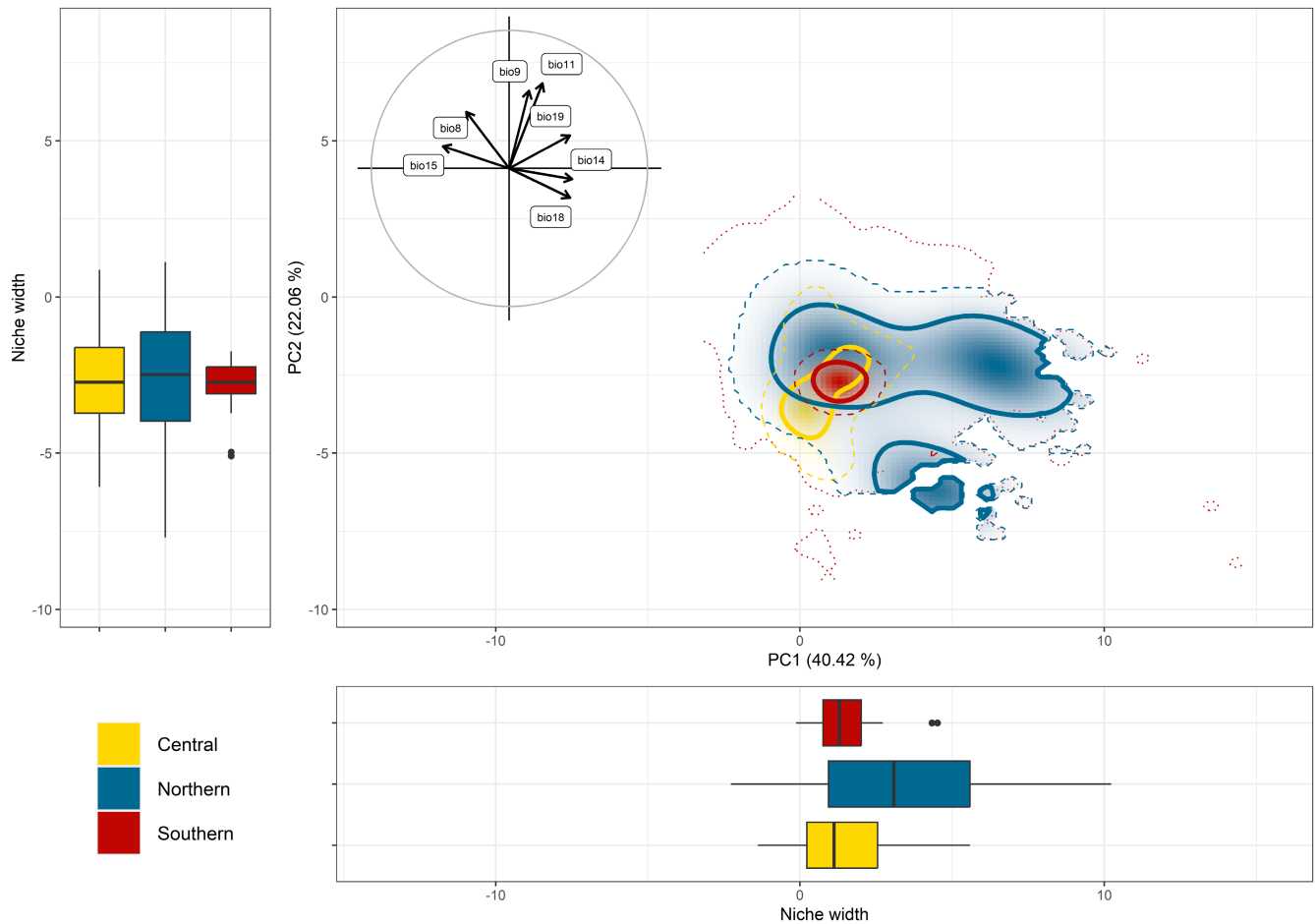
FIGURE 6 Gelada population niches along the first two principal component analysis (PCA) axes (PC1 and PC2) of bioclimatic variables recorded at occurrence sites and in the background environment within the study area. Red: niche calculated from central population sites; yellow: niche calculated from 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
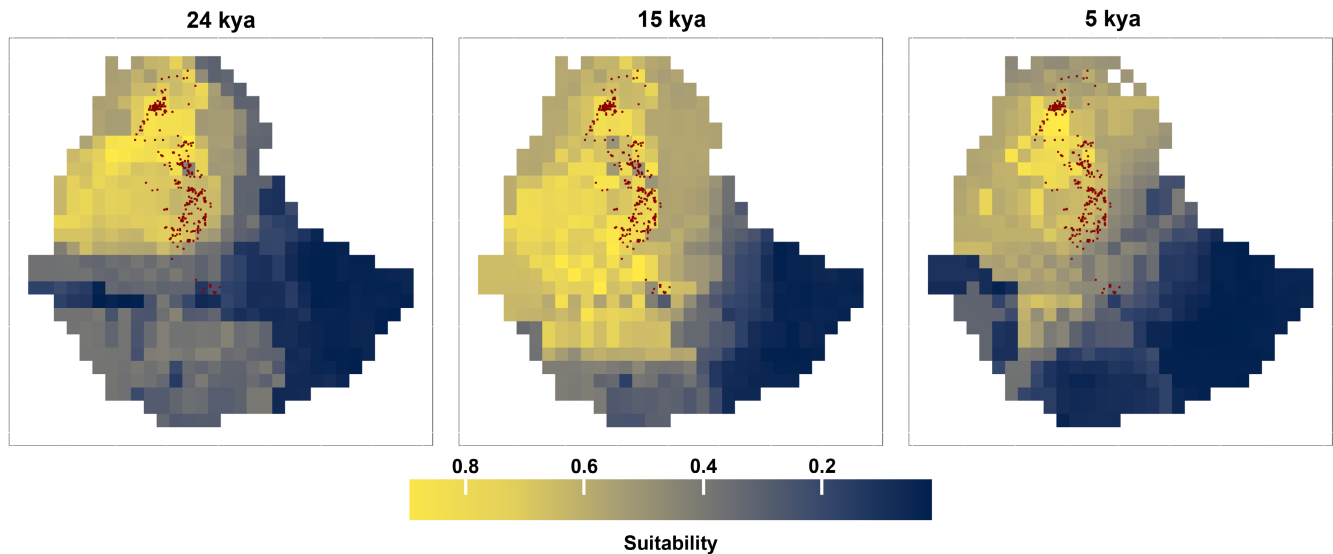
FIGURE 7 Predicted habitat suitability maps of *Theropithecus gelada* in Ethiopia at three different past time frames. Higher values indicate the areas that were more favourable for gelada habitation. Red dots represent the current geographical location of the gelada sites.

(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 TMRCAs, the Y chromosome being two times younger than the mitogenome (0.24 Mya vs. 0.49 Mya). 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. 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 (1600–2300 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.

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.

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., 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), giaa134. https://doi.org/10.1093/gigascience/giaa134

Belay, G., & Shotake, T. (1998). Blood protein variation of a new population of gelada baboons (*Theropithecus 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., Cappo, 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., Popinga, 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 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., … 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 ral 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 Bordallo, 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 theropiths: 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

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., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigyarto, 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 theropith 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, 12–4315.

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

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., 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

Haileselasie, 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 Leopoldio, Carolinae 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., & Jermiin, 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., Böhner, 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., 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*, 3–255. 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., 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

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., Cowlishaw, 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., Daugherity, 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., Rumpler, Y., Schneider, M. P., Silva, A., O'Brien, S. J., & Pecon-Slattery, 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 Abbyssinien 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-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., ... 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., Fontsere, 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., Nógues-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., Largen, M. J., & Kock, D. (1976). Catalogue of the mammals of Ethiopia. *Monitore Zoologico Italiano*, *8*, 1–118.

Yalden, D. W., Largen, 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