International Journal of Primatology (2025) 46:705–736
https://doi.org/10.1007/s10764-025-00494-2

Sex-mediated Gene Flow in Grayfoot Chacma Baboons (*Papio ursinus griseipes*) in Gorongosa National Park, Mozambique

Maria Joana Ferreira da Silva · Paula Tralma · Ivo Colmonero-Costeira · Mariana Cravo-Mota · Rassina Farassi · Philippa Hammond · Lynn Lewis-Bevan · Marion K. Bamford · Dora Biro · Tina Lüdecke · Jacinto Mathe · René Bobe · Cristian Capelli · Susana Carvalho · Felipe I. Martínez

Received: 25 October 2024 / Accepted: 3 April 2025 / Published online: 16 May 2025

Abstract

Dispersal behavior influences gene flow and the spatial distribution of genetic diversity, which is crucial for a species' evolutionary trajectory and population persistence under environmental changes. We used gene flow as a proxy to investigate dispersal patterns in the grayfoot chacma baboon (*Papio ursinus griseipes*) in Gorongosa National Park (GNP), central Mozambique. The baboons inhabit a mosaic landscape with a seasonally variable environment. Thirty-two years ago, GNP was the epicenter of a major war that severely reduced apex predators, resulting in limited mammalian predation on baboons. We aimed to characterize genetic diversity, examine the extent and direction of sex-biased gene flow at different time frames and investigate changes in population size and recent migration events. We collected 121 non-invasive DNA samples and analyzed uni- and bi-parentally inherited markers, comprising mitochondrial DNA, autosomal and Y-linked microsatellites, at two geographic locations (GNP and Catapú Forest Reserve) 150 km apart.

Handling Editor: Joanna M. Setchell

Maria Joana Ferreira da Silva and Paula Tralma are shared first co-authors.

Susana Carvalho and Felipe I. Martínez are shared senior authors.

Badges earned for open practices: Open Data. Experiment materials and data are available in the repository at GenBank (under Accession numbers PP398582–PP398684, and Dryad repository https://doi.org/10.5061/dryad.gb5mkkwxv).

We observed high genetic diversity and no evidence of a recent population decline. We identified six mitochondrial haplotypes, including a genetically distinct one in Catapu Forest Reserve. We found molecular evidence for historical and current male-mediated gene flow and female philopatry. Our results highlight the resilience of dispersal patterns in *Papio* sp. in diverse and seasonally variable ecosystems which have been disturbed by anthropogenic activities.

Keywords Male-biased dispersal · Southern Africa · Zambezi River · Warfare · Behavioral flexibility

Introduction

Dispersal, in which an individual moves away from their place of birth before breeding (e.g., natal dispersal, Pusey & Packer, 1987), is a key trait that affects groups and populations at various geographic scales. Firstly, dispersal is a core behavior mediating gene flow, and determines the spatial distribution of genetic diversity, which in turn defines a species' evolutionary trajectory and population persistence in the face of environmental change (Ferreira da Silva & Bruford, 2017). Secondly, the departure and inclusion of individuals in social units affects group composition and the nature of social interactions (Di Fiore, 2003). In mammals, dispersal tends to be sex mediated; males have a propensity to disperse, with females typically remaining in their natal group to reproduce (Pusey & Packer, 1987), although some species show the opposite trend (e.g., the Guinea baboon, *Papio papio*, Kopp et al., 2015, the chimpanzee, *Pan troglodytes*, Wilson, 2012, McCarthy et al., 2018, and the red colobus *Piliocolobus badius temminckii*, Minhós et al., 2013). The patterns of sex-biased dispersal observed in species are usually explained by the action of selective forces that are specific to each sex during the species' evolutionary history and dispersal patterns tend to be alike among closely related species (Jolly, 2020; Lawson Handley & Perrin, 2007).

Molecular methods have been widely used to investigate the extent and direction of dispersal because immigrants can be identified in the recipient population and traced back to their most probable origin (Di Fiore, 2009). Sex-biased dispersal over the historical timeframe of a population maintained in successive generations (i.e., "sex-bias gene-flow", Lawson Handley & Perrin, 2007, referring to two or more generations ago), is generally reflected in the genetic diversity, phylogeny, and spatial variation of uniparentally inherited markers, such as the mitochondrial DNA (mtDNA) and the Y-chromosome. More recent differences in dispersal between sexes (i.e., "instantaneous dispersal", Lawson Handley & Perrin, 2007, referring to up to one generation ago) can be detected using individual-based assignment methods, which are based on highly variable autosomal multi-locus microsatellite genotypes of post-dispersal and reproductively mature individuals (Lawson Handley & Perrin, 2007, Prugnolle & de Meeus, 2002).

Baboons (*Papio* sp.) are a group of ecologically and socially flexible primates distributed across sub-Saharan Africa and the Arabian Peninsula, with currently six

recognized species (Fischer et al., 2019). Their great behavioral flexibility (Swedell, 2011) may have evolved as a response to the highly variable paleoenvironments of Africa (the variability selection hypothesis, Alberts & Altmann, 2006). Studies of baboons report changes in home range size and use, foraging areas and effort, dietary habits, activity time-budget, and group composition (sex-ratio and size) in response to changes in food availability and perceived predation risk as well as to human activities (Altmann & Muruthi, 1988; Barton et al., 1992; Bronikowski & Altmann, 1996; Cowlishaw, 1997; Dunbar, 1992; Hill & Dunbar, 2002). Moreover, although sex-biased dispersal patterns tend to be conserved among baboon phylogenetic lineages (Jolly, 2020), variation in dispersal behavior at the population level has been suggested where natural predators are absent (the chacma baboon, *Papio ursinus*, Anderson, 1987) and in areas where baboons are hunted by local communities (the Guinea baboon, *Papio papio*, Ferreira da Silva et al., 2018). For instance, in a South African population of chacma baboons, a species in which dispersal is described to be male-biased (Fischer et al., 2019), a high number of females were observed to transfer between groups in an environment without predation and low ecological inter-individual competition (Anderson, 1987; Henzi et al., 2000).

Gorongosa National Park (GNP) in central Mozambique (Fig. 1) was once considered one of Africa's most diverse parks and populated by very large herds of herbivores and carnivores (Tinley, 1977). The region was a combat zone during the Mozambican war of independence against Portuguese occupation (1964–1977) and the Civil War (1979–1992) between the Mozambican National Resistance (Renamo) and the Liberation Front of Mozambique (Frelimo). During the Mozambican Civil War, which was considered one of the most violent and destructive in post-colonial African history, GNP was used as Renamo's headquarters (Seibert, 2003) and wildlife was hunted, which resulted in the elimination of 90–99% of all large herbivores and carnivores, including leopards (*Panthera pardus*), lions (*Panthera leo*) and spotted hyenas (*Crocuta crocuta*) (Hatton et al., 2001; Stalmans et al., 2019). Mammal communities are now recovering (Pringle, 2017; Stalmans et al., 2019).

Studies of the taxonomic identity and evolutionary history of baboons in GNP suggest that they are grayfoot chacma baboons (*Papio ursinus griseipes*), although they show considerable genetic proximity to yellow baboons (*Papio cynocephalus*) (Martinez et al., 2019; Santander et al., 2022). The impact of warfare and associated disturbances of the ecosystem on the inhabited GNP, including dispersal and gene flow, and the demographic trajectory of the population, remains poorly understood. The baboon population is thought to have undergone a significant reduction in the frequency of interactions with mammal predators (e.g., leopards, lions, and spotted hyenas) in the last two to three baboon generations (i.e., the last 20–30 years, based on a baboon generation time of 10 years, Altmann & Alberts, 2003; Bronikowski et al., 2002). While it is unknown whether the baboon's population size decreased during the war due to hunting, local people in the area ate baboon meat (Hatton et al., 2001) and are thought to still do so (M. Stalmans, personal communication, Fig. 1S, supplementary material).

We aimed to: i) characterize the distribution of genetic diversity in the GNP baboon population, ii) investigate the extent and direction of sex-mediated dispersal at different geographic scales and time periods, using gene flow as a proxy, iii)

Fig. 1 Location of the study site of *Papio ursinus griseipes* in Southern Africa and in central Mozambique, in relation to the Zambezi River and the city of Beira (data collected from 2017 to 2019). **a**) shows the range of subspecies of chacma (*Papio ursinus*) and yellow (*Papio cynocephalus*) and of Kinda baboons (*Papio kindae*). Distribution polygons are from the International Union for Conservation of Nature, IUCN. Male baboon illustrations copyright 2020 Stephen D. Nash/IUCN SSC Primate Specialist Group. Used with permission. **b**) Left: location of the two geographic populations studied—Gorongosa National Park (GNP) and Catapu Forest Reserve (CFR). Sites are distanced approximately 150 km apart. Right: location of fecal samples collected in GNP at the core of the protected area, near the Lake Urema (GNP core) and at a location closer to the western entrance of the park (BM). Figure also depicts the location of baboon groups estimated to live inside the park's limits (data by Stalmans et al., 2019). **c**) location of the five major habitats distinguished in GNP by Daskin et al., (2016) i) Midlands miombo woodland ii) Alluvial fan acacia, iii) Floodplain grasslands around lake Urema; (iv) colluvial fan savannas and (v) Cheringoma Plateau miombo woodlands and forested limestone gorges. The ii, iii and iv habitat types are within the Rift Valley.

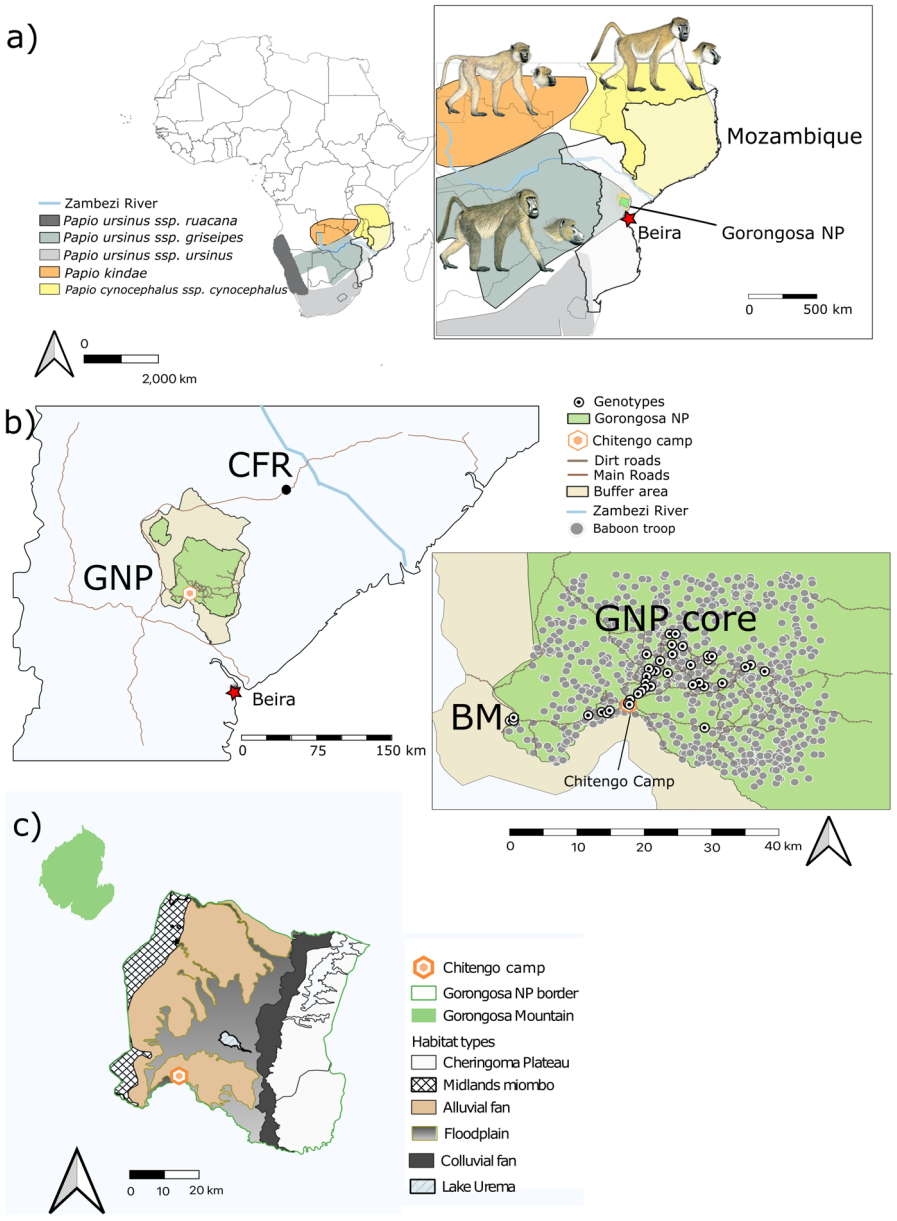
identify migrant baboon individuals in the park, and iv) investigate demographic changes in the population. We analyzed non-invasive DNA samples for uni- and bi-parentally inherited markers (mitochondrial DNA D-loop haplotypes, multilocus autosomal and Y-linked microsatellite data) collected across the region, including sites within GNP and the nearby area of Catapu Forest Reserve (CFR; Fig. 1b), and used population and individual-based statistical methods. We estimated genetic diversity and the population structure using mtDNA, Y-linked haplotypes and nuclear genetic markers and investigated "historical" (i.e., dispersal over several generations, referred to "sex-bias gene flow" in Lawson Handley & Perrin, 2007) corresponding to two or more baboon generations ago. We estimated "current dispersal" which refers to sex-specific gene flow patterns up to one generation ago (referred to as "instantaneous dispersal" in Lawson Handley & Perrin, 2007) by comparing the distribution of assignment indices and the spatial genetic autocorrelation between males and females (Goudet et al., 2002; Prugnolle & de Meeus, 2002).

We hypothesized that the grayfoot chacma in GNP shows a male-biased gene-flow pattern, as observed for other populations of chacma baboons (Fischer et al., 2019; Jolly, 2020). We predicted that: i) mtDNA haplotypes are more structured spatially than Y-linked haplotypes and nuclear genetic markers at the same spatial scale, and ii) when compared with females, males have lower mean probability of assignment to the population where they were sampled, and show significant genetic similarity at larger distances (Di Fiore, 2009; Melnick & Hoelzer, 1992; Prugnolle & de Meeus, 2002).

Methods

Study Site and Species

We carried out the study in GNP (18°43′56.54"S, 34°16′31.58"E, total area 3,770 km^2) and in the Catapú Forest Reserve (CFR, 16°50′54.11"S, 34°11′11.94"E, total area 250 km^2) (Fig. 1a and 1b). Both sites are in central Mozambique in the province of Sofala at the southern margin of the Zambezi River.



GNP is located between the Pungue and Nhandue rivers (Stalmans & Beilfuss, 2008) and at the southern end of the Great Rift Valley. GNP is a diverse and seasonally variable environment. The park comprises 15 different landscapes which are grouped into five major habitats: i) midlands miombo woodland, ii) alluvial fan acacias, combretum woodlands, and palm savannas, iii) open floodplain grasslands, iv) colluvial fan savannas, and v) Cheringoma Plateau miombo

woodlands and forested limestone gorges (Daskin et al., 2016; Stalmans & Beilfuss, 2008; Fig. 1c). The amount of rainfall varies across the year and across the park and about 40% of GNP core area floods every year (Stalmans & Beilfuss, 2008; Tinley, 1977).

CFR is a forestry concession located in the Cheringoma district on the main North/South highway (EN1), approximately 32 km from the southern margin of the Zambezi River and 150 km north from GNP core (Fig. 1b). CFR is classified floristically as Swahilian/Maputaland Regional Transition Zone. Baboons at CFR were reported to show a slightly different phenotype from those in GNP (Zinner et al., 2015).

DNA Sampling

We collected fecal samples from unidentified individual baboons opportunistically along unpaved roads and in areas usually frequented by baboons, such as the edge of drinking spots or foraging paths, during the dry seasons of 2017, 2018, and 2019. We also obtained one muscle sample from an individual who was the subject of infanticide, and one blood sample from another individual during GPS collar placement as part of a different study. Both tissue samples have been used to generate high and low coverage whole genome sequence data to investigate the evolutionary history of baboons in Gorongosa (Santander et al., 2022). At GNP, we mainly sampled two locations at a maximum distance of 30 km — around Lake Urema (GNP core) and at Boé Maria (BM) approximately 11 km from the closest sampling site in GNP core (Fig. 1b, left). At CFR, we sampled four sites 2–6 km apart (not shown). We geo-referenced fecal samples using a global positioning system (GPS) device (Garmin GPSMAP 64 s). We noted the sample's preservation status, the sex of the individual (when observed), and the number of baboons observed together at the time of sampling. We collected samples at least 2 m apart to minimize the likelihood of sampling the same individuals. We used gloves, facemasks, and hairnets during sampling to limit human cross-contamination. We preserved the samples until DNA extraction using the "two-step protocol" (Roeder et al., 2004), in which 5 ml of fecal material is collected from the outside of the sample, by scraping the surface using a wooden stick, and immediately immersed in 99% ethanol for 24–48 h. We then transferred the samples to a tube containing 30 g of Silica Gel (Type III, S- 7625, indicating for desiccation, Sigma–Aldrich) and stored until DNA extraction. We preserved the tissue samples in 99% ethanol at room temperature until DNA extraction.

DNA Extraction

We transported fecal samples to the Center for Research in Biodiversity and Genetic Resources laboratory, Porto University, Portugal, for DNA extraction. We extracted total genomic DNA of fecal samples using the QIAamp® DNA Stool Mini Kit (QIAGEN ®) with some modifications from the manufacturer's protocol (Ferreira da Silva et al., 2014). We eluted DNA extracts in 200 ul Buffer AE and stored them at –20ºC. We took several precautions to avoid contamination with exogenous

DNA during sample extraction (following Ferreira da Silva, 2012). We used negative controls in all extraction procedures to test for contamination with human DNA or cross-contamination between samples. We extracted DNA from the tissue samples in the molecular lab at Chitengo camp in GNP (Fig. 1b), using the QIAGEN DNeasy Blood & Tissue Kit following the manufactures' protocol. We eluted DNA extracts in 200 ul Buffer AE and stored them at –20ºC.

DNA Amplification and Data Production

We used the polymerase chain reaction (PCR) to recover 490 base pairs (bp) of the mitochondrial control region (hypervariable region I) (mtDNA) (Hapke et al., 2001), 14 autosomal microsatellite loci, and the Y chromosome linked DYS576 *locus* (supplementary material 1 and Table 1S, supplementary material 2). The microsatellite loci were developed in humans, and cross-amplify in *Papio* sp. (Ferreira da Silva et al., 2014). We chose the Y-locus (DYS576) because it is one of the few human-derived markers on the Y chromosome showing variation in *Papio* sp. (e.g., Jolly et al., 2011). We identified the sex of the individuals using the molecular protocol developed by Di Fiore (2005), in which a fragment of the amelogenin X gene (~ 200 bp) and a fragment of the SRY gene (~ 165 bp) are co-amplified in a multiplex PCR. We describe primer sequences and the PCR amplification cycling details in supplementary material 1 and 2. We used various procedures to limit cross-contamination between samples and from external DNA (supplementary material 3).

We sequenced mtDNA fragments bi-directionally using Sanger technology. We checked sequences and polymorphic positions using Geneious® v. 4.8.5 (Kearse et al., 2012). We created a consensus using both the forward and reverse sequences of the amplified fragments. We searched the NCBI database (https://www.ncbi.nlm.nih.gov/) for the most similar voucher using the BLAST algorithm to confirm species identity. We aligned sequences using the Geneious® v. 4.8.5 in-built algorithm and trimmed them to the length of the shortest sequence.

We amplified the 14 autosomal microsatellite loci and the Y-chromosome locus using five multiplex PCR systems (Table S1, supplementary material 2). The range of amplified alleles size varies between approximately 123 and 270 bp. All markers were tetranucleotide repeats, except D7S503 and D5S1457 that were dinucleotides (Table S1, supplementary material 2). We analyzed successful PCR products using Centro Testagem Molecular (CTM, at CIBIO Portugal) fragment analysis service and run them on an ABI3730XL capillary analyzer using a 16 GeneScanTM 500 LIZ ® size-standard.

We genotyped microsatellite loci following a modified version of the "multi-tubes" approach (Taberlet et al., 1996). We estimated the number of required repetitions across loci necessary to obtain reliable consensus genotypes using empirical data, following Ferreira da Silva et al., (2014) (supplementary material 4). As a result, we carried out a minimum of four amplifications per locus per sample. We quantified the reliability of consensus genotypes using the mean Quality Index (QI, Miquel et al., 2006) across loci. We removed genotypes with a QI below 0.55 from the final dataset (Miquel et al., 2006). We estimated the probability of identity (PI),

which is the probability that two individuals sampled randomly from the population have the same genotype at all typed loci, and the PI between siblings (*PIsibs*) (Waits et al., 2001) using GenAlEx v.6.503 (Peakall & Smouse, 2006). We estimated the minimum number of loci required to distinguish between different individuals using the study's set of microsatellite loci using poppr 2.9.3 package (Kamvar et al., 2015) in R v4.2.2 (R Core Team, 2022) coupled with RStudio v.1.3.1093 (Posit Team, 2023). We removed samples with the same genotype (possibly belonging to re-sampled individuals) from the dataset.

We tested for departure from Hardy–Weinberg equilibrium (HWE) per locus for the overall dataset and separately for GNP + BM and CFR using PopGenReport 2.2.2 R package (Adamack & Gruber, 2014). We corrected the *p*-value for significant deviations from HWE expectations using the Bonferroni adjustment for multiple comparisons. We used Micro-Checker version 2.2.3 (van Oosterhout et al., 2006) to test locus-specific deficiency in heterozygotes due to null alleles, stutter band-related scoring errors and large-allele dropout, all three with a 95% confidence interval. We computed an exact test for linkage disequilibrium (LD) between all pairs of loci for each sampling area using Arlequin v.3.11 (Excoffier & Lischer, 2015). We tested the significance of the association between pairs of loci via a likelihood ratio test using 10,000 permutations. We estimated the percentage of missing data for the overall dataset and by sampling locations using poppr 2.9.3 R package.

Statistical Analyses

Genetic Diversity

We computed mitochondrial DNA diversity summary statistics using DNAsp v5 (Librado & Rozas, 2009) and estimated the number of haplotypes, haplotype diversity (Hd; Nei, 1987), nucleotide diversity (π; Nei, 1987) and respective standard deviation (SD), and the mean number of pairwise differences (k) for each sampling regions and the overall dataset.

To describe the genetic diversity estimated using unique multi-locus autosomal microsatellite loci genotypes we used — the mean number of alleles per locus (*Na*), the number of equally frequent alleles (*Ne*), the observed heterozygosity (*Ho*), the unbiased expected heterozygosity (*UHe*) and the fixation index (*F*). We also calculated the allelic richness (*AR*) corrected for sample size by a rarefication method calculated in PopGenReport 2.2.2 R package (Adamack & Gruber, 2014) in R studio v.1.3.1093. We summarize the allelic patterns for the overall dataset and for each sampling site by the mean and standard error across loci of the *Na*, the *Na* with frequency over 5%, the *Ne*, the information index (I), the number of private alleles and the *He*, all calculated using GenAlEx v. 6.503.

Population Structure and Sex-mediated Gene Flow

We investigated population structure using population-based and individual-based methods. To investigate sex-biased dispersal for "historical" time periods (i.e., older than the current baboon generation), we compared genetic differentiation between the populations as estimated using maternal and bi-parental inherited markers. Considering a 10-year generation time based on life-history data from yellow baboons from Amboseli, Kenya (Altmann & Alberts, 2003) and anubis baboons (*Papio anubis*) from Gombe National Park, Tanzania (Bronikowski et al., 2002), we considered "historical" dispersal patterns (or "sex-bias gene-flow" in Lawson Handley & Perrin, 2007, i.e., two or more generations ago), as referring to periods before the 2000 s, which encompass the duration of the Mozambican Civil War.

We used Arlequin v3.11 (Excoffier & Lischer, 2015) to compute *Pairwise Fst* based on mtDNA D-loop haplotype frequencies and pairwise differences and a two-hierarchical level analysis of molecular variance (AMOVA). AMOVA estimates the total genetic variation explained by geographic populations membership (Excoffier et al., 1992). We calculated significance using 10,000 permutations. We used PopART (http://popart.otago.ac.nz) to construct a median joining network (Bandelt et al., 1999) with Epsilon and reticulation set to zero. We assessed spatially induced differentiation by distance-based RDA (db-RDA; Legendre & Anderson, 1999) to accommodate genetic distances between mtDNA sequences. RDA is an asymmetric canonical analysis which combines an ordination method and multiple regressions of independent predictor variables. When coupled with trend-surface analysis, it allows the fitting of linear gradients of genetic variance or more complex patterns over the geographical space (Legendre & Legendre, 2012). To calculate genetic distance matrices, we estimated the best-fit model of molecular evolution using ModelFinder (Kalyaanamoorthy et al., 2017). We limited the best-fit model search to the available models in ape 5.7.1 R package (Paradis & Schliep 2019). We selected the model K2P based on the Bayesian Information criteria (BIC). We used a third-degree orthogonal polynomial trend-surface of the geographic coordinates of sequences (long, lat, long * lat, $long^2$, lat^2, $long^2$ * lat, long * lat^2, $long^3$, lat^3) as predictors of the genetic differentiation among GNP and CFR (Legendre & Legendre, 2012). To prevent a possible over-fitting of the multiple regression models, we used a forward selection procedure with the following stopping criteria: variable significance of $p < 0.01$ and the adjusted R^2 (Blanchet et al., 2008). Subsequently, highly collinear variables (variance inflation factor, VIF > 5) were removed. After the selection procedure, we only retained the variable "long" as an explanatory geographic variable. We used vegan v2.6.4 R package (Oksanen et al., 2017) in R v4.2.2 (R CoreTeam 2022) coupled with RStudio v2023.06.2 +561 (Posit Team 2023). To obtain a visual representation of the main geographical structures, the fitted site scores of the first canonical axis were extrapolated using the inverse distance weighting (IDW) with power equal to two using gstat v2.1.1 package (Gräler et al., 2016; Pebesma, 2004).

We visualized population structure using autosomal microsatellites using a Principal Coordinate Analysis (PCoA) built using PopGenReport 2.2.2 R package. We tested for a relation between genetic and geographic differences using

Mantel tests in GenAlEx v. 6.503. The geographic Euclidean distances between genotypes were calculated using the Universal Transverse Mercator coordinate geographic system from each sample and converted to km in GenAlEx. The significance of the correlation was assessed using 9,999 random permutations. We estimated the contribution of geographic locations on genetic differentiation by combining redundancy (RDA) and trend–surface analyses for the whole dataset and for males and females separately. In RDA, we combined a principal components analysis (PCA) of the microsatellite loci genotypes data matrix and multiple regressions of the polynomial trend–surface of the geographic locations (Legendre & Legendre, 2012) following a similar procedure as described for dbRDA using mtDNA sequences. We retained "long", "long^2", and "lat*long^2 "for the whole dataset, the "long" variable for the male dataset and "long^3" and "lat*long" for the female dataset. We obtained statistical significance of the multiple regression models and each of the resulting canonical axes was by ANOVA-like permutation tests (9,999 permutations).

We ran the individual-based Bayesian STRUCTURE v2.1 software (Pritchard et al., 2000) using the"admixture model"(with parameters set to default), assuming correlated allele frequencies and the"LOCPRIOR model"(Hubisz et al., 2009). The use of the LOCPRIOR model allows detection of population structure in cases of gradual changes of allele frequencies across space and should not bias results to uncover significant structure when it is inexistent (Hubisz et al., 2009). We explored the number of clusters (K) between 1 and 10, running five independent repetitions for each K. A burn-in period of 100,000 steps was followed by 1,000,000 Markov Chain Monte Carlo (MCMC) runs.

A bias in STRUCTURE results may occur when the sampling of populations is highly uneven and, in such cases, individuals of smaller sample sizes may be erroneously assigned to a single cluster (Wang, 2017). Since sampling of the two geographic populations was unbalanced (GNP = 105 and CFR = 16 genotypes), we re-ran STRUCTURE using the above-mentioned parameters but choosing the alternative ancestry prior (an initial value of ALFA 1/K of 0.5 and inferred individual ALPHA for each population, following Wang, 2017). To test the effect of the presence of related individuals in STRUCTURE results, we followed Minhós et al., (2013); we estimated pairwise relatedness for the whole dataset using the Queller & Goodnight estimator (QG) in GenAlEx v. 6.503 (Queller & Goodnight, 1989), we removed one sample from dyads of an estimated 0.5 QG relatedness coefficient and re-ran STRUCTURE using parameters described above.

We identified the most likely number of clusters K using the posterior probability of K (following method to calculate the posterior probability described in Pritchard et al., 2000), and the highest estimated log-likelihood [Ln P(X/K)] and the ad hoc statistic ΔK developed by Evanno et al. (2005), both processed using STRUCTURE HARVESTER v6.8 (Earl & von Holdt, 2012). We used CLUMPPvs1.1.2 to permute and match individual-based outputs of independent runs (Jakobsson & Rosenberg, 2007). Graphs and plots were produced using POPHELPER online (Francis, 2017). To identify the threshold of assignment of individuals to a genetic cluster, we plotted the ranked q-values of each individual to each K (Beaumont et al., 2001). We detected a gap by visually inspecting the distributions of mean q-values across runs at 0.91, and

assigned individuals to a genetic cluster using the threshold of q > 0.90: the individuals with q value between 0.1 and 0.9 were considered to be a product of admixture.

We carried out a two-hierarchical level AMOVA (Excoffier et al., 1992) in Arlequin v.3.11 (Excoffier & Lischer, 2015). We performed a locus-by-locus AMOVA based on the number of different alleles as recommended for genetic datasets with missing data (Excoffier & Lischer, 2015). The results of the global AMOVA are an average (mean) weighted over loci. We tested the significance of covariance components and fixation indices using 16,000 permutations.

To investigate "instantaneous" sex-biased dispersal (i.e., up to one baboon generation ago, as used in Lawson Handley & Perrin, 2007), we compared the distribution of assignment indices and the spatial genetic autocorrelation between males and females (Goudet et al., 2002; Prugnolle & de Meeus, 2002). We could not carry out analyses reliant on grouping samples according to their respective social units of origin (such as the comparison between relatedness between males and females as in Prugnolle & de Meeus, 2002) because we collected fecal samples from unidentified individuals which were not assigned to a social unit prior to sampling. Thus, it is possible that we sampled and included pre-dispersal individuals (infants and juveniles) in our dataset. However, since these individuals represent a random combination of alleles not yet spatially distributed by sex-biased dispersal, their erroneous addition to the analyses testing "instantaneous" sex-mediated gene flow should erase, not augment, any sex-specific patterns of population structure (Prugnolle & de Meeus, 2002). We limited analyses to a maximum Euclidean distance of 40 km for samples collected in GNP.

We estimated the assignment index correction (*AIc*) (Favre et al., 1997), defined as the probability of assignment of an individual to the population where it was sampled using GenAIEX. We obtained a plot of the mean *AIc* for males versus females and a plot of the frequency distribution of *AIc* for the males and the females. The genetic signal of sex-biased dispersal is indicated when there is a significant difference in the frequency distribution of *AIc* values among males and females. Negative *AIc* values will characterize individuals with a higher probability of being immigrants. We tested significance in differences in *mAIc* between males and females by a non-parametric U-test implemented in GenAIEX "sex bias" module. As required, we only used samples without missing data ($N = 75$ unique genotypes, 41 males and 34 females).

We calculated the global autocorrelation coefficient r for males and females separately across distance classes. The spatial autocorrelation coefficient r measures genetic similarity ($r > 0$) and genetic dissimilarity ($r < 0$) between pairs of individuals separated in space by pre-defined geographical distances (Smouse & Peakall, 1999). We obtained the significance of r by a permutation process which defines an upper and lower bound of a 95% confidence interval. The observed r is considered significant when it lies outside the 95% confidence interval. Since we sampled randomly in space and not at regular distance intervals, we proceed to choose the width of the distance class by testing for different combinations of geographic distances and number of distance classes (Fontanillas et al., 2004). For that, we started by grouping genotypes in 5-km classes and subsequently conducted analyses at 5-km increments. We used this approach to understand the influence in identifying

significant spatial autocorrelation when increasing the distance class size width and changing the number of samples per distance class. We present the results of analyses in which genotypes are grouped in four distance classes 10 km wide (0–10, > 10–20, > 20–30, > 30–40 km, defined at the start point of the distance class) since we found this to be the best option in the trade-off between spatial resolution and sample size per distance class. We computed the test of significant heterogeneity between the overall patterns displayed by males and females using the "advanced multiple populations option" (9,999 permutations, 10,000 bootstraps), which also tests for significant differences between populations for each distance class (Smouse et al., 2008). The heterogeneity test is significant when $p < 0.01$ (Banks & Peakall, 2012). We used 105 unique genotypes (56 males and 49 females).

Identification of Migrants

To identify putative migrants, we ran STRUCTURE with the option USEPOPINFO to compute the probability that an individual belonged to either the GNP or CFR populations. We did not consider BM as an independent population in these analyses as sample size is below ten individuals. We ran the analysis three times, assuming migration rate (MIGPRIOR) to vary between 0.001 and 0.05 and GENSBACK to vary between 0 and 3. Burn-in and MCMC length were the same as described above, in the previous section. GENECLASS version 2.0 (Piry et al., 2004) was used to detect F0 immigrants (*i* first generation migrants) by estimating two likelihood-based test statistics: L_h (more appropriate when the source population might not have been sampled) and L_h/L_{max} (mostly used when all source populations were presumably sampled) (Paetkau et al., 2004). We identified critical values distinguishing between residents and immigrants using the Rannala and Mountain (1997) Bayesian method and the Paetkau et al. (2004) resampling Monte Carlo algorithm (1,000 simulations and an alpha level of 0.01). We compiled results from the two analyses and classified individuals as putative migrants when highlighted by at least one of the methods.

Inferring Demographic Changes

We used the mtDNA dataset and neutrality tests to evaluate significant departures from the expectations of a neutral model with constant population size (Ramírez-Soriano et al., 2008). Significant deviations suggest the disruption of one or more assumptions of the model, which can be caused by selective and demographic events changing population size. An increase of the population size or a selective sweep is commonly reflected in longer external branches in the population genealogy, which results in an excess of recent mutations (e.g., star-shaped genealogies); on the contrary, the decrease of the population size or balancing selection may result in an excess of old over recent mutations, and originate longer internal branches

(Ramírez-Soriano et al., 2008). We computed the D* and F* test statistics (Fu & Li, 1993), which only require intraspecific data, and estimated the critical values to determine statistical significance using DNAsp v5 (Librado & Rozas, 2009). Fu and Li's D* and F* tests compare the number of derived singleton mutations and the total number of derived nucleotide variants (D*) and the pairwise differences between sequences (F*). We also calculated the D test statistic proposed by Tajima (1989) and the respective confidence intervals (a two-tailed test), assuming a beta distribution of the statistic. Tajima's D compares estimates of the number of segregating sites and the mean pairwise difference between sequences. Tajima's D tends to be negative for genealogies that are approximately star-like and positive for the ones with deep bifurcations (Ramírez-Soriano et al., 2008). We computed neutrality tests for the overall dataset, for each sampled geographic population and for CFR, removing the three individuals with the most divergent haplotype (Hap_6), accounting for the possibility of a recent introduction of the haplotype in the population by migrant individuals.

We used BOTTLENECK v.1.2.02 (Piry et al., 1999) to test for population size reduction using microsatellite data. We used the stepwise (SMM) and two-phase mutation models (TPM) (Luikart & Cornuet, 1998) for datasets of samples grouped by geographic populations and for the overall dataset. The TPM model comprised 95% single step mutations and 5% multi-step mutations. Variance for mutation size was set to 12 (Piry et al., 1999). We ran 100,000 simulations. We tested differences between the expected heterozygosity levels under mutation-drift equilibrium (HetEQ) and the observed heterozygosity using the Wilcoxon's sign rank test. We also assessed whether the distribution of allele frequencies was shifted (i.e., the mode-shift test) or normal *(i.e.,* L-shaped).

Ethical Note

The research complied with rules and protocols approved by GNP and adhered to the legal requirements of Mozambique, including genetic benefit sharing (Nagoya Protocol). All except two samples were obtained non-invasively from unidentified individuals, without manipulation or perturbation of their daily behavior. We collected one muscle tissue sample opportunistically from an animal already deceased and one blood sample during GPS collaring initiatives for other studies. In the latter case, the animal was handled only by the park veterinarian to collect blood. The blood collection and GPS collaring were approved by the ACER Committee at the University of Oxford, UK (ref APA/1/5/ACER/23 Jan2018). GNP authorized collection and exportation of fecal samples and DNA extracts (Permits PNG/DSCi/C80/2017, PNG/DSCi/C117/2018, PNG/DSCi/R/249/2022).

Data Availability The datasets generated during the current study are available in the GenBank (under Accession numbers PP398582–PP398684, and Dryad repository https://doi.org/10.5061/dryad.gb5mkkwxv).

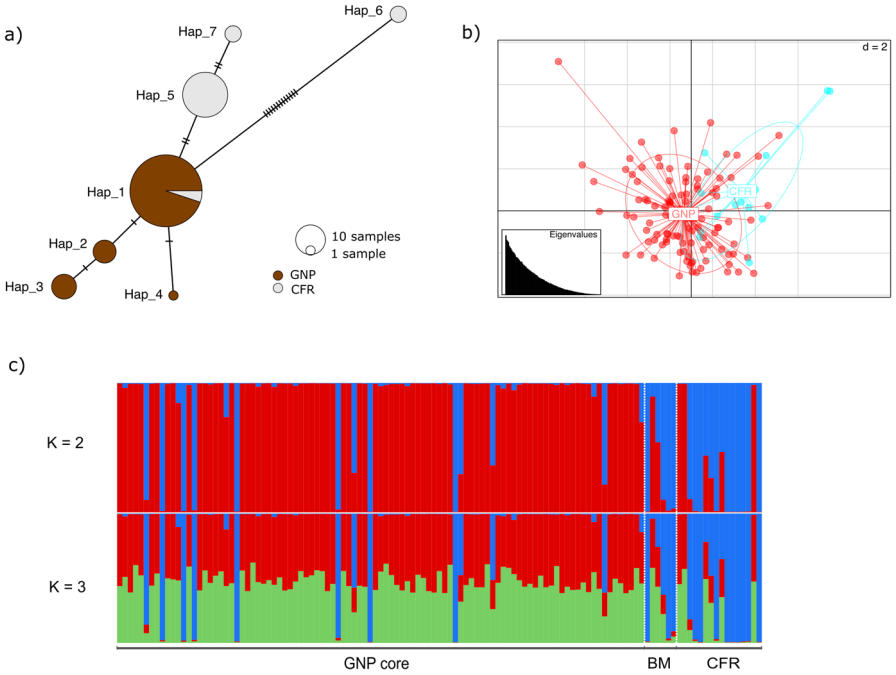


Fig. 2 Population structure as estimated using mitochondrial DNA (mtDNA) and autosomal data of *Papio ursinus griseipes* in Gorongosa National Park (GNP) and Catapu Forest Reserve (CFR) in central Mozambique (genetic data collected from 2017 to 2019). **a**) Median joining network of D-loop mtDNA haplotypes built using PopART (Epsilon and reticulation were set to zero). The seven haplotypes distinguished were arbitrarily named by numbers in the figure (from 1 to 7). The circles sizes are proportional to number of individuals harbouring the haplotype and the number of dashes in branches represent the number of mutations that separate the haplotypes. Colours indicate sampling sites (brown – GNP and light grey – CFR). Hap_6 (harbored by three individuals) differs 13 substitutions from the most similar haplotype (Hap_1). **b**) PcoA plot based in autosome microsatellite loci built using PopGenReport 2.2.2 R package. The first two coordinates of the PcoA explained 11.21% of the total variation. **c**) STRUCTURE results using an alternative ancestry prior. The STRUCTURE plots show the q membership of individuals for the K = 2 and K = 3 clustering solutions. The geographic location for each individual is indicated below the figure; GNP core ($N_{GNP\ core}$ = 99 genotypes), BM – Boé Maria (N_{BM} = 6 genotypes), and CFR—Catapú Forest Reserve (N_{CFR} = 16 genotypes). Plots were produced using POPHELPER online.

Results

DNA Sampling and DNA Extraction and Amplification Success

We extracted whole genomic DNA from 260 fecal samples, one muscle tissue and one blood sample. We obtained forward and reverse mitochondrial (mtDNA) D-loop region consensus sequences based on high-quality chromatograms for 102 different individuals (based on unique multi-locus genotypes, see below). We obtained 70 mtDNA sequences for GNP (including GNP core and BM) and 32 for CFR. The multi-locus genotypes for 15 of the sequences from CFR had low reliability

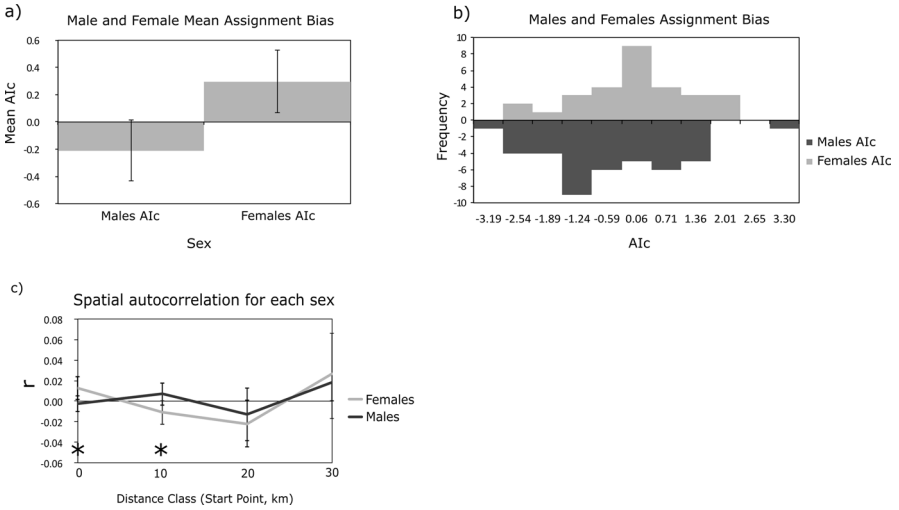


Fig. 3 "Instantaneous" sex-biased dispersal assessed for a maximum Euclidean distance of 40 km for samples from *Papio ursinus griseipes* population of Gorongosa National Park (GNP) in central Mozambique (genetic data collected from 2017 to 2019). **a**) plot of the mean Assignment Index correction (*AIc*) for males and females. We used N = 75 unique genotypes (41 males and 34 females) without missing data; **b**) a plot of the frequency distribution of *AIc* for the males and females; **c**) the correlogram of global autocorrelation coefficient *r* calculated for males and females separately across four distances classes. We used N = 105 unique multi-locus genotypes (56 males and 49 females). Significant differences between the sexes are emphasised by*.

(QI < 0.53), so we did not include them in the microsatellite loci dataset used for analyses.

The D-loop sequences showed a high percentage of identity (95.82% to 96.94%) with *Papio ursinus* GenBank vouchers JX964908.1, JX964906.1, JX964908.1, JN116783.1, JX946205.2, and JX964908.1, so we assumed the samples were from chacma baboons.

The Sex Determination Protocol Distinguished 59 Females and 65 Males.

We obtained 121 unique genotypes with a QI above 0.55 (mean QI = 0.79, 4.25% missing data, Fig. 2S supplementary material 5). This includes 105 genotypes for GNP (99 from GNP core; six from BM) and 16 genotypes for CFR (one from site 1; two from site 2; five from site 3; eight from site 4). The 14 microsatellite loci discriminated between unrelated individuals using two loci and related individuals with six loci (pID = 4.3×10^{-17} and pIDsibs = 1.09×10^{-6}). The genotype accumulation curve reached a plateau with four loci, suggesting the ability to discriminate between different individuals is reached with five or fewer loci (Fig. 3S, supplementary material 5). We removed five repeated genotypes from within 500 m of one another in GNP from the final dataset because they may belong to the same individual. We found no evidence for null alleles, stuttering errors or large allelic dropout for the overall dataset or when we split genotypes by localities. Four loci (D7S503, D5S1457, D6S501, D10S611) were not in HWE when we pooled all genotypes, but none remained significant after adjusting for multiple comparisons (using the

Bonferroni adjustment of $p < 0.0037$). We found evidence of LD ($p < 0.05$) when we analyzed samples by region (GNP 12 pairs and CFR 13 pairs) and one pair in GNP (D12S375 and D3S1766) remained significant after Bonferroni adjustment ($p < 0.0037$). We interpreted these results as indicating population substructure due to the likely presence of related individuals in the GNP sample and retained the loci in the analyses.

We obtained reliable genotypes for the DYS576 locus for 21 of the 65 individuals distinguished as males.

Genetic Variation

The 433 bp fragment of the mtDNA control region contained 18 polymorphic sites (18 substitutions) and five parsimony-informative sites (two variants). We distinguished seven haplotypes (or 17 haplotypes if we consider alignment gaps). Most haplotypes were one or two mutations apart from the others (average k = 1.99) except for Hap_6. Haplotype 6 was present in three individuals in CFR, and differs by 13 mutational steps (substitutions) from the most similar haplotype (Hap_1). CFR had slightly larger nucleotide diversity than GNP (Table I).

The autosomal microsatellite loci showed high levels of genetic diversity (Table 2S, supplementary material 6). The mean number of alleles (*Na*) over the 14 microsatellite loci was 8.9 (SE 0.65), ranging between five (D1S533) and 13 (D10S611) (Table 2S, supplementary material 6). The genetic diversity and allelic patterns are similar across sampling sites (Table I, Fig. 4S, supplementary material 6). The allelic richness corrected for sample size was 5.4 for GNP and 5.3 for CFR (Table I). We found six alleles for the locus DYS576 across the 21 males genotyped. DYS576 allele size ranged between 284 and 306 bp (Table I).

Population Structure

The distribution of the D-loop mtDNA haplotypes was spatially structured by sampling site, and most haplotypes are site-specific (Fig. 2a). The most frequent haplotype, Hap_1, present in 58% (59/102) of sequences obtained, was much more prevalent in GNP (56/59, 95%) than in CFR (3/59 = 5%). The geographical location of sampling regions explained 18.14% of the total mtDNA variance ($r^2 = 0.17$, ANOVA-like $F_{(1,\ 101)} = 20{,}71$, $p < 0.001$). Pairwise *Fst* as inferred from mtDNA D-loop haplotype frequencies (*Fst* haplotype frequencies = 0.57) and pairwise differences (*Fst* pairwise differences = 0.52) were significant between sampling sites ($p < 0.001$, significance tests based on 16,000 permutations, Table 4S, supplementary material 7). The AMOVA based on haplotype frequencies revealed that a comparable proportion of variation was present between (55.9%) and within sampling regions (44.1%, Table 5S, supplementary material 7), and the fixation index (*Fst* = 0.56) was significantly different from zero. RDA suggested that 18.14% of the variance of the genetic differentiation could be explained by the geographic location of haplotypes (Fig. 5Sa, supplementary material 7).

The PCoA of the autosomal microsatellite loci did not suggest genetic dissimilarity among GNP and CFR (Fig. 2b). The first two coordinates of the PCoA explained

Table I Genetic diversity of *Papio ursinus griseipes* in Gorongosa National Park and Catapu Forest Reserve, in Central Mozambique estimated using non-invasive DNA samples, collected between 2017 and 2019

Region	Autosomal microsatellite loci							mtDNA D-loop					DYS576	
	N	Hexp	UHe	Ho	Na	AR	F	N	nH	Hd	π	k	N	Alleles (N)
GNP	105	0.778 ± 0.021	0.782 ± 0.021	0.778 ± 0.022	8.64 ± 0.57	5.42	− 0.001 ± 0.013	70	4	0.347 ± 0.068	.12% ± 0.0002	0.518	17	294 (7) 298 (5) 302 (2) 306 (2)
CFR	16	0.767 ± 0.020	0.798 ± 0.021	0.798 ± 0.041	6.5 ± 0.43	5.29	0.033 ± 0.044	32	4	0.472 ± 0.1	.79% ± 0.002	3.399	4	284 (2) 290 (1) 294 (1)
Overall	121	0.773 ± 0.014	0.790 ± 0.015	0.761 ± 0.023	7.57 ± 0.41		0.016 ± 0.023	102	7	0.611 ± 0.0019	.48% ± 0.00098	1.994	21	

Genetic diversity is shown per sampling region (GNP – Gorongosa National Park and CFP – Catapú Forest Reserve) and for the overall genetic dataset. We estimated genetic diversity using two types of genetic markers (14 autosomal microsatellite loci and the Y-linked locus DYS576, and a fragment in mitochondrial D-loop region). Autosomal microsatellite loci—Number of unique genotypes (N), Expected Heterozygosity (Hexp), Unbiased expected heterozygosity ± SD (UHe), observed heterozygosity ± SD (Ho), mean number of alleles (Na), Allelic Richness (AR) based on 10 individuals, Fixation Index (F). mtDNA D-loop – Number of sequences (N); Number of haplotypes (nH); Haplotype diversity (Hd) and Nucleotide diversity (π) ± Standard Deviation and mean number of nucleotide differences (k). DYS576 – Number of genotypes (N), observed alleles in base pairs (Alleles) and number of males (N) displaying the allele

11.21% of the total genetic variation. We found a significant positive linear relationship between genetic and geographic distances (Mantel test: Rxy = 0.342, R^2 = 0.11, $p < 0.001$). Likewise, RDA suggested a significant but small contribution of geographic location on genetic differentiation (Fig. 5S, supplementary material 7). For the whole dataset, the polynomial trend-surface of the geographic coordinates explained 3.34% of the total variance ($r^2 = 0.017$, ANOVA-like $F_{(2, 118)} = 2.06$, $p < 0.001$, Fig. 5S b, supplementary material 7). Re-running RDA for the two sexes separately yielded similar results, although the geographic location of populations explained a slightly higher proportion of total variance in females when compared to males (females: 6.48% of explained variance, $r^2 = 0.010$, ANOVA-like $F_{(2, 50)} = 1.73$, $p < 0.001$; males: 2.67% of explained variance, $r^2 = 0.010$, ANOVA-like $F_{(1, 58)} = 1.59$, $p < 0.01$, Fig. 5S c and d, supplementary material 7).

STRUCTURE runs suggested that the sampled individuals formed one single genetic unit. While K = 6 was the highest modal value in the ΔK distribution (Fig. 6S supplementary material 7), the clustering solution with the largest Log(likelihood) and posterior probability across all runs was $K = 1$ (posterior probability $K1 = 0.97$). When $K = 2$, the individuals have an equal assignment probability to the two clusters. In contrast, in STRUCTURE analyses run with an alternative prior (ALPHA inferred for each population and initial ALPHA of 0.5), the clustering solution with the highest modal value in the ΔK distribution and largest Log(likelihood) was for $K = 2$ (Fig. 7S, supplementary material 7). The individuals sampled in GNP were mostly assigned to cluster 1 (82%) and 10% were assigned to cluster 2 (Fig. 2c). At CFR, ten of the 16 individuals were assigned to cluster 2, but three individuals were assigned to cluster 1 (Fig. 2c). We considered the remaining three individuals to be admixed between clusters and left them unassigned.

We found 16 dyads with an estimated Queller & Goodnight relatedness coefficient above 0.5. After removing one sample from each dyad and re-running STRU CTURE using the same parameters as previous runs, we found no difference in clustering results (one single genetic unit, results not shown), suggesting that relatedness between individuals did not affect STRUCTURE runs significantly.

Pairwise *Fst* values estimated using microsatellite loci suggested low differentiation between the two sampled localities ($Fst = 0.017$, $p < 0.05$ Table 4S, supplementary material 7). Locus-by-locus AMOVA averaged over the 14 loci revealed greater variation within sampling locations (97.4%) than among GNP and CFR (2.61%). Sites show low genetic differentiation (average *Fst* over all loci = 0.026) (Table 5S, supplementary material 7).

The most frequent allele of the Y-linked locus DYS576 (allele 294, found in eight of 21 males) was shared between GNP and CFR (Table I).

Assignment Tests

mAIc was negative for males (*mAIc* males = –0.271 standard error (SE) = 0.229 and confidence interval (CI) = 0.448) and positive for females (*mAIc* females = 0.327, SE 0.221, CI 0.433) in GNP (genotypes with no missing data; 34 identified as

females and 41 as males) (Fig. 3a). *mAIc* differed significantly between males and females for the upper tail (U-test, $p = 0.037$), with females displaying more positive *AIc* values than males in the frequency distribution (Fig. 3b). The most negative *AIc* was for a male (male sample BF30, Fig. 3b).

Spatial Autocorrelation

The total spatial autocorrelation pattern was not significantly different between males and females in the core area of GNP (total Omega for data = 7.552, $p = 0.057$) but the two sexes showed significant differences in spatial genetic structure at shorter distances (up to 20 km). From 0 to 10 km, female baboons showed significantly higher spatial autocorrelation coefficient than males (r $\text{Females}_{10\text{ km}} = 0.012$ and $\text{Males}_{10\text{ km}} = -0.002$, $p = 0.038$). The spatial autocorrelation coefficient was significantly lower and negative between 10 and 20 km for females than for males (r $\text{Females}_{20\text{ km}} =$—0.011 and r $\text{Males}_{20\text{ km}} = 0.007$, $p = 0.031$) (Fig. 3c).

Identification of Migrants

STRUCTURE analyses identified two females sampled in GNP as having a significant probability of being migrants (GENSBACK = 1 and MIGRPRIOR = 0.05) and none as having migrant ancestry (GENSBACK = 2 and MIGRPRIOR = 0.05) (Table 6S, supplementary material 8). When we repeated the analysis with a close-to-zero restrictive migration rate (MIGPRIOR = 0.01), no individual was highlighted as a migrant or having migrant ancestry. One female sampled in GNP and identified as first-generation migrant was assigned to the genetic unit with the lowest frequency in GNP in STRUCTURE runs with an alternative prior (i.e., cluster 2) and the other female was considered admixed (Table 6S, supplementary material 8). Analyses carried out using GENECLASS detected seven individuals as "first-generation migrants" using the likelihood ratio L_{home}/L_{max} and alpha 0.01. Two of these individuals identified as "first-generation migrants" were sampled in GNP and five in CFR. Of these putative migrants, we determined the sex of two samples in GNP (two females) and two samples in CFR (one male and one female). For the three remaining putative first-generation migrants, we could not identify the sex of the individuals (Table 6S, supplementary material 8). STRUCTURE and GENECLASS analyses concur in the individuals identified as first-generation migrants in GNP (Table 6S, supplementary material 8). The male with most negative *AIc* (male sample BF30) was not identified as a putative migrant by these analyses.

Inferring Demographic Changes

We did not find deviations from a neutral model with constant population size for the overall dataset or for the GNP population separately. For CFR dataset, we computed positive, and significant Fu and Li's D* (D* = 1.59, p = 0.02) and Fu's Fs (Fs = 4.987, p = 0.022) statistics (Table 7S, supplementary material 9). When we

repeated the analyses without the three individuals with the most divergent haplotype in the CFR population (Hap_6), Fu and Li's D* and Fu's Fs remained positive but were not significant (D* = 1.062, $p > 0.10$, Fs = 0.66, $p = 0.284$).

We did not find evidence for a recent bottleneck using BOTTLENECK software in the overall microsatellite dataset or for the two geographic populations since we found no significant heterozygosity excess for the two models tested (e.g., overall dataset, Wilcoxon sign rank test two-tailed probability: TPM, $p = 0.19$; SMM, $p = 0.54$) and the distribution of allele frequencies was normal L-shaped.

Discussion

This study found high levels of genetic diversity and molecular evidence for male-biased gene flow in grayfoot chacma baboons in central Mozambique. The population shows higher levels of genetic diversity than in other chacma baboons, as estimated using microsatellite loci (i.e., GNP and CFR grayfoot chacma, *He* = 0.79 ± SD 0.015, mean number of alleles = 7.57 ± 0.41; Zambian grayfoot chacma *He* = 0.56–0.72 (Burrell, 2008); Tsaobis, Namibia chacma baboons: *He* = 0.63 (Huchard et al., 2010); Fig. 8S, supplementary material 10). These differences are probably not the result of ascertainment bias from the selection of microsatellite loci used here (Amos & Harwood, 1998) since another study using whole genomic data from two grayfoot chacma baboons sampled in GNP found a similar pattern of higher genetic variation when compared to other available chacma baboon genomes (reported SNP heterozygosity: GNP baboons = 0.12; chacma of unknown origin = 0.06, Santander et al., 2022). Moreover, the population at GNP is phenotypically very variable (Martinez et al., 2019). Individuals may display one or more features considered characteristic of chacma, yellow and Kinda baboons (*Papio kindae*) (i.e., chacma: robust male body build, bent tail carriage, or downwardly flexed facial orientation; yellow and Kinda baboons: fur color, lightly colored ventral hair, pink infraorbital skin) in a mosaic-like pattern (Fig. 3 in Martinez et al., 2019). This phenotypic variation is likely to result from ancient introgression by male chacmas into yellow baboon populations and nuclear swamping, which is thought to be the demographic event at the origin of the *P. u. griseipes* subspecies (Keller et al., 2010; Santander et al., 2022; Sithaldeen et al., 2015; Sørensen et al., 2023; Zinner et al., 2009).

Alternatively, the phenotypic variation observed in the GNP population may be related to more recent secondary contact and gene flow between grayfoot chacma, yellow, or Kinda baboons (Santander et al., 2022) possibly caused by changes in the drainage course of the Zambezi River (e.g., Nugent, 1990). Gene flow between species of the genus *Papio* is not uncommon (Sørensen et al., 2023). Although Sørensen et al. (2023) describe an admixture area with a phenotypically homogenous population, individuals in admixture zones frequently show phenotypic traits resembling more than one baboon species and admixed populations display increased genetic variation when compared to neighboring ones of parental species (e.g., Charpentier et al., 2012; Chiou et al., 2021; Jolly et al., 2011; Sørensen et al., 2023).

The high autosomal genetic diversity in the GNP grayfoot chacma baboons may also be interpreted as the result of a prolonged period of large effective population

size. We did not find evidence for a significant contraction of the population in the last two to three generations (no significant differences between HetEQ and the observed heterozygosity). Thus, the recent demographic history of baboons in GNP does not seem to match what was recorded for other mammals in the region, whose populations were severely reduced by hunting during and after the Civil War (Hatton et al., 2001; Stalmans et al., 2019). Although it is thought that baboons were hunted for meat during the Mozambican Civil War, chacma baboons have not been identified as a particularly preferred species (Hatton et al., 2001). Baboons were probably targeted less than other mammals (e.g., elephant *Loxodonta africana*, buffalo *Syncerus caffer sp.* or hippopotamus *Hippopotamus amphibius*, Hatton et al., 2001) because of their relatively smaller body size and consequent lower profit from the trade of their carcasses. The positive relation between body size and preference by hunters and the gradual replacement for smaller body-size species when larger ones became rare has been observed in other locations in Africa of intense hunting during and after armed conflicts (e.g., Braga-Pereira et al., 2020; Ferreira da Silva et al., 2021).

Alternatively, an increase in the baboon population size after the Civil War ended (in 1992) might have contributed to maintaining the high genetic diversity we detected. It has been hypothesized that the GNP baboon population has grown due to very low predation pressure, and high availability of food, coupled with limited wildlife hunting under the management of the Gorongosa Restoration Project (Stalmans et al., 2019). Baboons' natural predators — leopards, lions and spotted hyenas — were present in large numbers when the park was first established (Tinley, 1977), but were greatly reduced or, in the case of leopards and spotted hyenas, extirpated during the civil war due to intense hunting (Hatton et al., 2001). The current populations of carnivores are far below the size of those reported in pre-war times (Stalmans et al., 2019). The baboons are thought to have rarely been in contact with mammalian predators during the day in the last two or three generations and display behaviors suggestive of low levels of interaction with carnivores, such as unusually high terrestrial activity during crepuscular and nocturnal hours, including sleeping on the ground (Hammond et al., 2022, Hammond et al., accepted). A growth of the baboon population size reported within the protected area (Stalmans et al., 2019) is in line with the demographic pattern of other species lacking significant carnivore predation (e.g., waterbuck, *Kobus ellipsiprymnus* in Stalmans et al., 2019). Nevertheless, the increase in group number described in Stalmans et al., (2019) can also be explained by permanent group fission, which would have been promoted by an increase in group size and the need to reduce local resource competition or by decrease in predation pressure (Ramos-Fernandez & Aureli, 2018).

In contrast to a high autosomal genetic diversity, the level of mtDNA D-loop genetic diversity we found (N = 102, Hd = 0.61, $\prod$ = 0.0048 SE 0.00098) is lower than that estimated for the whole range of grayfoot chacma (N = 72 sequences, 31 haplotypes, Hd = 0.96 and $\prod$ = 0.0011 SE 0.05) and Cape chacma baboons (*P. u. ursinus*, N = 57 sequences, 30 haplotypes, Hd = 0.96 and $\prod$ = 0.0031 SE 0.06) using the same mtDNA region (Table I in Sithaldeen et al., 2015). Lower levels of mtDNA diversity for the Mozambican grayfoot chacma baboon (Hd ~ 0.5 and $\prod$ < 0.005) may be the result of a historical bottleneck or a founder event by few females/

lineages and restricted female-mediated gene flow in male-biased dispersal species. The combination of the two factors is predicted to result in low levels of haplotype diversity locally (Di Fiore, 2009). For instance, this effect was described in groups of grayfoot chacma in Zambia, in which only one haplotype was found per deme (Burrell, 2008).

We found a very distinct haplotype (Hap_6) in CFR, which was the underlying cause for higher levels of nucleotide diversity in that population when compared to GNP. The finding of a distinct haplotype in CFR suggests secondary contact between differentiated lineages in that area, which may be the result of sampling recent immigrants to the population, possibly males originating from other populations. However, we were unable to test whether the individuals harboring Hap_6 were recent migrants to the CFR population because we only obtained partial genotypes of low QI for the three individuals. Moreover, the results of neutrality tests to infer historical changes in population size seem to have been affected by the presence of Hap_6 in CFR. The neutrality tests suggested different demographic histories among the sampled regions; whereas we found no significant deviation from the expectations of a constant size at mutation-drift equilibrium for GNP, we found positive and significant Fu and Li's D* and Fu's Fs for CFR, which suggest an excess of ancestral haplotypes in that population, possibly the result of population size contraction. However, when we re-ran the analyses removing Hap_6 from the CFR dataset, all statistics were non-significant, suggesting that the result of population size contraction was related to the presence of Hap_6 in CFR. Future work could assign the most probable geographic origin of Hap_6 in Southern Africa using available datasets (e.g., Sithaldeen et al., 2015) and use approximate Bayesian computation to test for the most probable demographic scenarios of the population (Sousa et al., 2012).

We detected seven putative first-generation migrants out of 121 sampled individuals (6%, both males and females). However, the proportion of immigrants in the population is likely to be lower than this estimate because some of the individuals identified as putative immigrants may be pre-dispersal and sub-adults and descendants of true immigrants, which we were unable to identify and exclude from our database. Although the populations inhabiting the buffer zone or occurring outside the park have not been investigated, baboons are expected to occur in lower densities outside the park's core area (i.e., around lake Urema) since they are targeted for hunting and trapping (Lindsey & Bento, 2012, Fig. 1S, supplementary material) and may be killed by communities in retaliation for crop-foraging (Easter et al., 2019). Our team did not observe baboons in the buffer zone outside the park limits, where most human infrastructures are located (2018 and 2019, MJFS pers. obs.) and successful sampling outside the park's limits was only possible in CFR (150 km away from GNP), where baboons were previously reported to be present *(e.g.,* Zinner et al., 2015). Thus, our results suggest that the increase in the number of baboon groups recorded by aerial surveys reported by Stalmans et al., (2019) is unlikely to have been caused by a substantial influx of immigrants towards the park's core.

The CFR site shows signs of increased immigration rate when compared to GNP. We detected a relatively larger number of putative recent migrants (five individuals in CFR vs two in GNP), and a very distinctive haplotype (Hap_6), which may have

originated from a geographically distant population. We also found a larger number of alleles for DYS576 (three alleles in a total of four males genotyped), suggestive of male gene flow towards CFR. Unlike GNP, CFR is close to the Zambezi River hydrographical basin (about 32 km, Fig. 1b), which could be used as an ecological corridor by baboons from distant populations (Vale et al., 2015), possibly all the way north to Zambia where hybridization between chacma and yellow baboons has been reported (Jolly et al., 2011). Interestingly, baboons at CFR occur in the range of *Papio cynocephalus strepitus* (in Fig. S1 in Kopp et al., 2023) and studies report a larger resemblance to yellow baboons in their general hair color than in the baboons inhabiting core areas of the GNP (Zinner et al., 2015). Further studies could investigate the origin of immigrants reaching CFR and the implications of gene flow for the phenotypic variation of baboons observed in the region. The analysis of a larger set of Y chromosome microsatellite loci is expected to generate informative haplotypes and enable a more detailed reconstruction of the relationships among males within and across populations (Mutti et al., 2023).

The significant structure of mtDNA variation across 150 km, is similar to the strong regional geographic structure of D-loop mitochondrial lineages found at the species range scale (Sithaldeen et al., 2015). In contrast, we could not detect significant population structure for the whole spatial scale of the study using autosomal markers (e.g., we found no pattern of total autosomal genetic variation in the PCoA and STRUCTURE runs suggested one single genetic unit). Nevertheless, STRUCTURE analyses using an alternative prior to compensate for unequal sampling effort suggested the presence of two genetic units, broadly dividing samples from GNP and CFR (Fig. 2). The analysis also identified individuals sampled in GNP which could be assigned to the most frequent genetic unit in CFR site with a high probability and vice versa, overall suggesting ongoing gene flow (i.e., up to one generation ago) between the two geographic populations.

We found molecular evidence for "historical" male-biased dispersal and female philopatry for the grayfoot chacma baboon across 150 km (e.g., dispersal over several generations, Lawson Handley & Perrin, 2007). Specifically, we found significant substructure in mtDNA variation and very little for autosomal markers, which is the pattern expected for male-biased dispersing and female philopatric species (Di Fiore et al., 2009). Additionally, the most frequent allele of Y-linked locus was shared between the two localities, which also suggests male gene flow across 150 km. We also found that most of mtDNA genetic variation is distributed among GNP and CFR (56%), while almost all autosomal variation occurred within sites (97.4%). Our results align with the genetic patterns of other chacma baboons' populations in Southern Africa in which males disperse and females remain in the natal groups (e.g., Zambia — grayfoot chacma: proportion of mtDNA variation among regions = 75.65% vs autosomal variation among regions = 2.2%; yellow: among regions mtDNA = 100% vs autosomal = 2.0%, and Kinda: among regions mtDNA = 94.8% vs autosomal = 3.3%, Burrell, 2008). Although mtDNA variation among populations in our study is lower than in other male-biased dispersal baboons (74–100%, Burrell, 2008), this is probably the consequence of different sampling strategies between studies, such as the inability to sample geographically delimited social units and use demes as the unit in analyses. Nevertheless, mtDNA variation among regions

in Mozambique is still very different to what was found for baboons, in which gene-flow is mediated mostly by females, in particular when the comparison is carried out for similar geographic distances (e.g., mtDNA variation among regions ~150 km apart, the Guinea baboon = 11.36% and hamadryas baboon = 33.48%, Kopp et al., 2014, the Guinea baboon in ~150 km in Guinea-Bissau: among regions mtDNA variation = 8.7% and autosomal genetic variation = 3.34%, Ferreira da Silva, 2012).

Within GNP, we found fine-scale genetic structure and differences between the sexes. Females show significant genetic similarity in distances up to 10 km (positive genetic autocorrelation, Peakall et al., 2003) which should correspond to the presence of matrilineal groups expected for chacma baboons. Moreover, females tend to be genetically more dissimilar between 10 and 20 km, which could be interpreted as limited female-mediated gene flow at that distance. Males, in contrast, tend to be similar between 10 and 20 km, suggesting dispersal and reproduction to up to 20 km. We found no significant differences between the sexes for larger distance classes. A weak but significant correlation in Mantel tests support the observation that individuals are genetically more similar at shorter distances. These results suggest "instantaneous" dispersal (e.g., in the present baboon generation) is male biased at small distances in GNP (i.e., < 20 km).

Conclusions

We found molecular evidence for "historical" and "instantaneous" male-biased gene flow for the grayfoot chacma baboon population in central Mozambique. This study adds empirical evidence to the literature showing that sex-biased dispersal patterns tend to be conserved among different populations in cercopithecine monkeys facing different environments and levels of seasonality (Di Fiore & Rendall, 1994; e.g., deforestation, McCarthy et al., 2018). Our study also emphasizes selection for dispersal and philopatry behaviors in baboons, as hypothesized by Jolly (2020). Sex-mediated dispersal patterns are stable within *Papio* species, possibly because individuals that diverge from the norm may incur fitness disadvantages (Jolly, 2020). Extending our understanding of dispersal patterns of *Papio* sp. is important since evolutionary forces shaping genetic and phenotypic diversity are thought to be related to male-mediated gene flow (for instance, introgressive hybridization and the "nuclear swamping" hypothesis) (Jolly, 2020; Keller et al., 2010; Sørensen et al., 2023). Thus, future studies assessing intra-species behavioral variation in dispersal and the local environmental factors impacting gene flow may provide important insights into the evolutionary history of baboons.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10764-025-00494-2.

Acknowledgements We are very grateful to Gorongosa National Park, the Park Warden Pedro Muagura, the fiscais who provide support, and the Park staff, including Vasco Galante and Patricia Alvares da Guerra. We appreciate the help with logistics provided by Miguel Lajas, Jason Denlinger, and

Tongai Castigo. We thank Catapu Forest Reserve manager James Ant White. We thank students from the Paleo-Primate Field School who took part in aspects of field research. We thank CTM/CIBIO lab staff — Susana Lopes, Patricia Ribeiro, and Diana Castro — for their outstanding help and support. We thank Laura Macamo and the staff from the molecular lab in Gorongosa National Park. Research and export permits were granted by Gorongosa National Park (Permits PNG/DSCi/C80/2017, PNG/DSCi/C117/2018, PNG/DSCi/R/249/2022), with the support of Dr Marc Stalmans, and the Park Warden Mr. Pedro Muagura. MJFS worked under an associate researcher contract (Fundação para a Ciência e Tecnologia, Portugal https://doi.org/10.54499/CEECIND/01937/2017/CP1423/CT0010). ICC and MM were supported by a FCT PhD scholarship (ICC: https://doi.org/10.54499/SFRH/BD/146509/2019); MM: PD/BD/140228/2018 and COVID/BD/152649/2022). S.C. (NGS- 57285R), T.L. (NGS-51478R-18), and R.B. (NGS- 51140R- 18) were supported by grants by the National Geographic Society. M.B. acknowledges funding from PAST (Palaeontological Scientific Trust, South Africa). F.I.M. from VRI-UC and from ANID/FONDECYT 1240657, T.L. from the Deutsche Forschungsgemeinschaft (LU 2199/2- 1), D.B. from the Jackie Lambert Research Fund of St Hugh's College Oxford and CC by the COMP-R Initiative, funded by the 'Departments of Excellence' program of the Italian Ministry for University and Research (MUR, 2023-2027). This study was supported by the Gorongosa Restoration Project, by ICArEHB (Interdisciplinary Center for Archaeology and Evolution of Human Behaviour) funded by the Portuguese Foundation for Science and Technology (FCT) under program UIDP/04211/2020, and by ANID/FONDECYT 1240657. We thank the editor and two anonymous reviewers for attentive comments which greatly improved this manuscript.

Authors contributions Conceptualization: MJFS, SC, CP and FIM.
Data curation: MJFS, PT, ICC, JM, RF, PH, LLB, RB and MCM.
Investigation: MJFS, PT, CP, FIM, JC, PH, LLB, MB and RF.
Formal analyses: MJFS, PT, ICC, LLB, PH and MCM.
Project administration: SC, CC, FIM, RB.
Funding acquisition: TL, SC, RB, CC, MB, FIM.
Writing – original draft and review and editing: all authors.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Inclusion and diversity One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ + community. One or more of the authors of this paper self-identifies as living with a disability. The author list includes contributors from the location where the research was conducted, who participated in study conception, study design, data collection, analysis, and/or interpretation of the findings.

References

Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution, 5*(4), 384–387. https://doi.org/10.1111/2041-210X.12158

Alberts, S. C., & Altmann, J. (2006). The evolutionary past and the research future: Environmental variation and life history flexibility in a primate lineage. In L. S. S. Leigh (Ed.), *Reproduction and Fitness in Baboons: Behavioral, Ecological and Life History Perspectives*. Springer.

Altmann, J., & Alberts, C. (2003). Variability in reproductive success viewed from a life-history perspective in baboons. *American Journal of Human Biology, 15*, 401–409. https://doi.org/10.1002/ajhb.10157

Altmann, J., & Muruthi, P. (1988). Differences in daily life between semiprovisioned and wild-feeding baboons. *American Journal of Primatology, 15*, 213–221.

Amos, W., & Harwood, J. (1998). Factors affecting levels of genetic diversity in natural populations. *Phil. Trans. r. Soc. Lond. B, 353*(1366), 177–186. https://doi.org/10.1098/rstb.1998.0200

Anderson, C. M. (1987). Female transfer in baboons. *American Journal of Physical Anthropology, 73*(2), 241–250. https://doi.org/10.1002/ajpa.1330730211

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Banks, S. C., & Peakall, R. (2012). Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Molecular Ecology, 21*, 2092–2105. https://doi.org/10.1111/j.1365-294X.2012.05485.x

Barton, R. A., Whiten, A., Strum, S. C., Byrne, R. W., & Simpson, A. J. (1992). Habitat use and resource availability in baboons. *Animal Behaviour, 43*(5), 831–844. https://doi.org/10.1016/S0003-3472(05)80206-4

Beaumont, M., Barratt, E. M., Gottelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K., & Bruford, M. W. (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology, 10*, 319–336. https://doi.org/10.1046/j.1365-294x.2001.01196.x

Braga-Pereira, F., Peres, C. A., Campos-Silva, J. V., Santos, C.V.-D., & Alves, R. R. N. (2020). Warfare-induced mammal population declines in Southwestern Africa are mediated by species life history, habitat type and hunter preferences. *Scientific Reports, 10*, 15428.

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal, 8*(1), 204–218.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology, 89*(9), 2623–2632. https://doi.org/10.1890/07-0986.1

Bronikowski, A. M., & Altmann, J. (1996). Foraging in a variable environment: weather patterns and the behavioral ecology of baboons. *Behavioral Ecology and Sociobiology*, *39*(1), 11–25. http://www.jstor.org/stable/4601229

Bronikowski, A. M., Alberts, S. C., Altmann, J., Packer, C., Dee Carey, K., & Tatar, M. (2002). The aging baboon: Comparative demography in a non-human primate. *PNAS, 99*(14), 9591–9595. https://doi.org/10.1073/pnas.142675599

Burrell, A. (2008). *Phylogenetics and population genetics of Central African baboons*. PhD thesis. New York University.

Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., . . . Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. *Molecular Ecology*, *21*(3), 715–731. https://doi.org/10.1111/j.1365-294X.2011.05302.x

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

Cowlishaw, G. U. Y. (1997). Trade-offs between foraging and predation risk determine habitat use in a desert baboon population. *Animal Behaviour, 53*(4), 667–686. https://doi.org/10.1006/anbe.1996.0298

Daskin, J. H., Stalmans, M., & Pringle, R. M. (2016). Ecological legacies of civil war: 35-year increase in savanna tree cover following wholesale large-mammal declines. *Journal of Ecology, 104*(1), 79–89. https://doi.org/10.1111/1365-2745.12483

Di Fiore, A. (2003). Molecular genetic approaches to the study of primate behavior, social organization, and reproduction. *Am J Phys Anthropol, Suppl, 37*, 62–99. https://doi.org/10.1002/ajpa.10382

Di Fiore, A. (2005). A rapid genetic method for sex assignment in non-human primates. *Conservation Genetics*, *6*, 1053–1058 https://doi.org/10.1007/s10592-005-9086-5

Di Fiore, A., Link, A., Schmitt, C., & Spehar, S. (2009). Dispersal patterns in sympatric woolly and spider monkeys: Integrating molecular and observational data. *Behaviour, 146*(4–5), 437–470. https://doi.org/10.1163/156853909X426345

Di Fiore, A., & Rendall, D. (1994). Evolution of social organization: A reappraisal for primates by using phylogenetic methods. *Proceedings of the National Academy of Sciences of the United States of America., 91*(21), 9941–9945. https://doi.org/10.1073/pnas.91.21.9941

Di Fiore, A. (2009). Genetic approaches to the study of dispersal and kinship in New World primates. In P. A. Garber, A. Estrada, J. C. Bicca-Marques, E. W. Heymann, spsampsps K. B. Strier (Eds.), *South American Primates: Comparative Perspectives in the Study of Behavior, Ecology, and Conservation* (pp. 211–250). Springer. https://doi.org/10.1007/978-0-387-78705-3_9

Dunbar, R. I. M. (1992). Time: A hidden constraint on the behavioural ecology of baboons. *Behavioral Ecology and Sociobiology, 31*(1), 35–49. https://doi.org/10.1007/BF00167814

Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Easter, T., Bouley, P., & Carter, N. (2019). Opportunities for biodiversity conservation outside of Gorongosa National Park, Mozambique: A multispecies approach. *Biological Conservation, 232*, 217–227. https://doi.org/10.1016/j.biocon.2019.02.007

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology, 14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Lischer, H. (2015). *Arlequin ver 3.5. An Integrated Software Package for Population Genetics Data Analysis*. Computational and Molecular, Population Genetics Lab (CMPG), Institute of Ecology and Evolution, University of Berne, Switzerland and Swiss Institute of Bioinformatics, Switzerland. http://cmpg.unibe.ch/software/arlequin35

Excoffier, L., Smouse, P., & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics, 131*, 479–491.

Favre, F., Balloux, F., Goudet, J., & Perrin, N. (1997). Female-biased dispersal in the monogamous mammal Crocidura russula: Evidence from field data and microsatellite patterns. *Proceedings of the Royal Society of London B Biological Sciences, 264*, 127–132. https://doi.org/10.1098/rspb.1997.0019

Ferreira da Silva, M. (2012). Hunting pressure and the population genetic patterns and sex-mediated dispersal in the Guinea Baboon in Guinea-Bissau. PhD thesis, Cardiff University. Cardiff. UK

Ferreira da Silva, M. J., & Bruford, M. W. (2017). Genetics and Primate Conservation. In A. Fuentes (Ed.), *The International Encyclopedia of Primatology*. Wiley Online.

Ferreira da Silva, M. J., Godinho, R., Casanova, C., Minhós, T., Sá, R., & Bruford, M. W. (2014). Assessing the impact of hunting pressure on population structure of Guinea baboons (*Papio papio*) in Guinea-Bissau. *Conservation Genetics, 15*(6), 1339–1355. https://doi.org/10.1007/s10592-014-0621-0

Ferreira da Silva, M. J., Kopp, G., Casanova, C., Godinho, R., Minhós, T., Sá, R., . . . Bruford, M. W. (2018). Disrupted dispersal and its genetic consequences: Comparing protected and threatened baboon populations (*Papio papio*) in West Africa. *PLos ONE*, *3*(4), e0194189. https://doi.org/10.1371/journal.pone.0194189

Ferreira da Silva, M. J., Minhós, T., Sá, R., Casanova, C., & Bruford, M. W. (2021). A qualitative assessment of Guinea–Bissau's hunting history and culture and their implications for primate conservation. *African Primates, 15*, 1–18.

Fischer, J., Higham, J., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, J., . . . Zinner, D. (2019). The natural history of model organisms: Insights into the evolution of social systems and species from baboon studies. *eLife 8*(e50989), 1–16. https://doi.org/10.7554/eLife.50989

Fontanillas, P., Petit, E., & Perrin, N. (2004). Estimating sex-specific dispersal rates with autosomal markers in hierarchically structured populations. *Evolution, 58*(4), 886–894. https://doi.org/10.1111/j.0014-3820.2004.tb00420.x

Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources, 17*, 27–32. https://doi.org/10.1111/1755-0998.12509

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics, 133*(3), 693–709. https://doi.org/10.1093/genetics/133.3.693

Goudet, J., Perrin, N., & Waser, P. (2002). Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Molecular Ecology, 11*(6), 1103–1114. https://doi.org/10.1046/j.1365-294x.2002.01496.x

Hammond, P., Lewis-Bevan, L., Biro, D., & Carvalho, S. (2022). Risk perception and terrestriality in primates: A quasi-experiment through habituation of chacma baboons (*Papio ursinus*) in Gorongosa National Park. *Mozambique. American Journal of Biological Anthropology, 179*(1), 48–59. https://doi.org/10.1002/ajpa.24567

Hammond, P., Gaynor, K., Easter, T., Biro, D. & Carvalho, S. (accepted) Landscape-scale effects of season and risk on the terrestrial activity patterns of chacma baboons (*Papio ursinus*). *American Journal of Biological Anthropology*.

Hapke, A., Zinner, D., & Zischler, H. (2001). Mitochondrial DNA variation in Eritrean hamadryas baboons (Papio hamadryas hamadryas): Life history influences population genetic structure. *Behavioral Ecology and Sociobiology, 50*(6), 483–492. https://doi.org/10.1007/s002650100393

Hatton, J., Couto, M., & Oglethorpe, J. (2001). Biodiversity and war: A case study of Mozambique. In K. Sullivan (Ed.), *Biodiversity Support Program*.

Henzi, S. P., Barrett, L., Weingrill, A., Dixon, P., & Hill, R. (2000). Ruths amid the alien corn: Males and the translocation of female chacma baboons. *South African Journal of Science, 96*(2), 61–62.

Hill, R. A., & Dunbar, R. I. M. (2002). Climatic determinants of diet and foraging behaviour in baboons. *Evolutionary Ecology, 16*(6), 579–593. https://doi.org/10.1023/A:1021625003597

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources, 9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x. Epub 2009 Apr 1.

Huchard, E., Knapp, L. A., Wang, J., Raymond, M., & Cowlishaw, G. (2010). MHC, mate choice and heterozygote advantage in a wild social primate. *Molecular Ecology, 19*(12), 2545–2561. https://doi.org/10.1111/j.1365-294X.2010.04644.x

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics, 23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jolly, C. (2020). Philopatry at the frontier: A demographically driven scenario for the evolution of multilevel societies in baboons (Papio). *Journal of Human Evolution, 146*, 102819.

Jolly, C. J., Burrell, A. S., Phillips-Conroy, J., Bergey, C., & Rogers, J. (2011). Kinda baboons (*Papio kindae*) and Grayfoot Chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue River Valley. *Zambia. American Journal of Primatology, 73*, 291–303. https://doi.org/10.1002/ajp.20896

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. (2015). 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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keller, C., Roos, C., Groeneveld, L. F., Fischer, J., & Zinner, D. (2010). Introgressive hybridization in southern African baboons shapes patterns of mtDNA variation. *American Journal of Physical Anthropology, 142*(1), 125–136. https://doi.org/10.1002/ajpa.21209

Kopp, G., Fischer, J., Patzelt, A., Roos, C., & Zinner, D. (2015). Population genetic insights into the social organization of Guinea baboons (*Papio papio*): Evidence for female-biased dispersal. *American Journal of Primatology, 77*(8), 878–889. https://doi.org/10.1002/ajp.22415

Kopp, G. H., Ferreira da Silva, M. J., Fischer, J., Brito, J. C., Regnaut, S., Roos, C., & Zinner, D. (2014). The influence of social systems on patterns of mitochondrial DNA variation in baboons. *International Journal of Primatology, 35*(1), 210–225. https://doi.org/10.1007/s10764-013-9725-5

Kopp, G. H., Sithaldeen, R., Trede, F., Grathwol, F., Roos, C., & Zinner, D. (2023). A comprehensive overview of baboon phylogenetic history. *Genes*, *14*(3), 614. Retrieved from https://www.mdpi.com/2073-4425/14/3/614

Lawson Handley, L. J., & Perrin, N. (2007). Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology, 8*, 1559–1578. https://doi.org/10.1111/j.1365-294X.2006.03152.x

Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs, 69*(1), 1–24. https://doi.org/10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2

Legendre, P., & Legendre, L. (2012). *Numerical ecology* (third edition). Elsevier.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lindsey, P., & Bento, C. (2012). *Illegal Hunting and the Bushmeat Trade in Central Mozambique. A Case-study from Coutada 9, Manica Province* (pp. 64). Traffic East/Southern Africa, Harare.

Luikart, G., & Cornuet, J.-M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology, 12*, 228–237. https://doi.org/10.1111/j.1523-1739.1998.96388.x

McCarthy, M. S., Lester, J. D., Langergraber, K. E., Stanford, C. B., & Vigilant, L. (2018). Genetic analysis suggests dispersal among chimpanzees in a fragmented forest landscape in Uganda. *American Journal of Primatology, 80*, e22902. https://doi.org/10.1002/ajp.22902

Martinez, F. I., Capelli, C., Ferreira da Silva, M. J., Aldeias, V., Alemseged, Z., Archer, W., . . . Carvalho, S. (2019). A missing piece of the Papio puzzle: Gorongosa baboon phenostructure and intrageneric relationships. *Journal of Human Evolution*, *130*, 1–20. https://doi.org/10.1016/j.jhevol.2019.01.007

Melnick, D., & Hoelzer, G. (1992). Differences in male and female dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *International Journal of Primatology, 13*, 379–393.

Minhós, T., Nixon, E., Sousa, C., Vicente, L. M., da Silva, M. F., Sá, R., & Bruford, M. W. (2013). Genetic evidence for spatio-temporal changes in the dispersal patterns of two sympatric African colobine monkeys. *American Journal of Physical Anthropology, 150*(3), 464–474. https://doi.org/10.1002/ajpa.22223

Miquel, C., Bellemain, E., Poillot, C., Bessiére, J., Durand, A., & Taberlet, P. (2006). Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes, 6*, 985–988. https://doi.org/10.1111/j.1471-8286.2006.01413.x

Mutti, G., Oteo-Garcia, G., Caldon, M., Capelli, C., et al. (2023). Assessing the recovery of Y chromosome microsatellites with population genomic data using Papio and Theropithecus genomes. *Scientific Reports, 13*, 13839. https://doi.org/10.1038/s41598-023-40931-x

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Nugent, C. (1990). The Zambezi River: Tectonism, climatic change and drainage evolution. *Palaeogeography, Palaeoclimatology, Palaeoecology, 78*, 55–69.

Oksanen, F.J., et al. (2017) Vegan: community ecology package. R package Version 2.4–3. https://CRAN.R-project.org/package=vegan

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. *Molecular Ecology, 13*(1), 55–65. https://doi.org/10.1046/j.1365-294x.2004.02008.x

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Peakall, R., Ruibal, M., & Lindenmayer, D. B. (2003). Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, rattus fuscipes *Evolution*, *57*(5), 1182–1195. https://doi.org/10.1111/j.0014-3820.2003.tb00327.x

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences, 30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity, 95*(6), 536–539. https://doi.org/10.1093/jhered/esh074

Piry, S., Luikart, G., & Cornuet, J.-M. (1999). Computer note. BOTTLENECK: A computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity, 90*(4), 502–503. https://doi.org/10.1093/jhered/90.4.502

Posit Team. (2023). RStudio: *Integrated Development Environment for R. Posit Software.* http://www.posit.co/

Pringle, R. M. (2017). Upgrading protected areas to conserve wild biodiversity. *Nature, 546*(7656), 91–99. https://doi.org/10.1038/nature22902

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Prugnolle, F., & de Meeus, T. (2002). Inferring sex-biased dispersal from population genetic tools: A review. *Heredity, 88*, 161–165. https://doi.org/10.1038/sj.hdy.6800060

Pusey, A. E., & Packer, C. (1987). Dispersal and philopatry. In B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, & T. T. Struhsaker (Eds.), *Primate Societies*. University of Chicago Press.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution, 43*, 258–275.

Ramírez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F., & Navarro, A. (2008). Statistical power analysis of neutrality tests under demographic expansions, contractions, and bottlenecks with recombination. *Genetics, 179*(1), 555–567. https://doi.org/10.1534/genetics.107.083006

Ramos-Fernandez, G., & Aureli, F. (2018). Fission-Fusion. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1–8). Springer International.

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, 94*(17), 9197–9201. https://doi.org/10.1073/pnas.94.17.9197

R Core Team. (2022). R: *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,* Vienna. https://www.R-project.org

Roeder, A. D., Archer, F. I., Poinar, H. N., & Morin, P. A. (2004). A novel method for collection and preservation of faeces for genetic studies. *Molecular Ecology Notes, 4*(4), 761–764. https://doi.org/10.1111/j.1471-8286.2004.00737.x

Santander, C., Molinaro, L., Mutti, G., Martínez, F. I., Mathe, J., Ferreira da Silva, M. J., . . . 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

Seibert, G. (2003). The vagaries of violence and power in post-colonial Mozambique. In *Rethinking Resistance: Revolt And Violence In African History* (pp. 254–276). Retrieved from https://hdl.handle.net/1887/12889

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

Smouse, P. E., Peakall, R., & Gonzales, E. (2008). A heterogeneity test for fine-scale genetic structure. *Molecular Ecology, 17*(14), 3389–3400. https://doi.org/10.1111/j.1365-294X.2008.03839.x

Smouse, P. E., & Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity, 82*(5), 561–573. https://doi.org/10.1038/sj.hdy.6885180

Sørensen, E. F., Harris, R. A., Zhang, L., Raveendran, M., Kuderna, L. F. K., Walker, J. A., . . . Rogers, J. (2023). Genome-wide coancestry reveals details of ancient and recent male-driven reticulation in baboons. *Science*, *380*(6648), eabn8153. https://doi.org/10.1126/science.abn8153

Sousa, V. C., Beaumont, M. A., Fernandes, P., Coelho, M. M., & Chikhi, L. (2012). Population divergence with or without admixture: Selecting models using an ABC approach. *Heredity, 108*(5), 521–530. https://doi.org/10.1038/hdy.2011.116

Stalmans, M., & Beilfuss, R. (2008). *Landscapes of the Gorongosa National Park*. Gorongosa Restoration Project. https://www.researchgate.net/publication/314878798_Landscapes_of_the_Gorongosa_National_Park

Stalmans, M. E., Massad, T. J., Peel, M. J. S., Tarnita, C. E., & Pringle, R. M. (2019). War-induced collapse and asymmetric recovery of large-mammal populations in Gorongosa National Park. *Mozambique. Plos ONE, 14*(3), e0212864. https://doi.org/10.1371/journal.pone.0212864

Swedell, L. (2011). African papionins: Diversity of social organization and ecological flexibility. In C. Campbell, A. Fuentes, K. MacKinnon, S. Bearder, & R. Stumpf (Eds.), *Primates in Perspective* (2nd ed., pp. 241–277). Oxford University Press.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., . . . Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194. https://doi.org/10.1093/nar/24.16.3189

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tinley, K. L. (1977). *Framework of the Gorongosa ecosystem*. D.SC. (Wildlife management), University of Pretoria, South Africa.

Vale, C., Ferreira da Silva, M., Campos, J., Torres, J., & Brito, J. (2015). Applying species distribution modelling to the conservation of an ecologically plastic species (*Papio papio*) across biogeographic regions in West Africa. *Journal for Nature Conservation, 27*, 26–36. https://doi.org/10.1016/j.jnc.2015.06.004

van Oosterhout, C., Weetman, D., & Hutchinson, W. F. (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes, 6*(1), 225–256. https://doi.org/10.1111/j.1471-8286.2005.01082.x

Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Molecular Ecology, 10*(1), 249–256. https://doi.org/10.1046/j.1365-294x.2001.01185.x

Wilson, M. L. (2012). Long-term studies of the chimpanzees of Gombe National Park, Tanzania. In P. Kappeler & D. Watts (Eds.), *Long-term field studies of primates* (pp. 357–384). Springer.

Wang, J. (2017). The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources, 17*(5), 981–990. https://doi.org/10.1111/1755-0998.12650

Zinner, D., Groeneveld, L. F., Keller, C., & Roos, C. (2009). Mitochondrial phylogeography of baboons (Papio spp.) – indication for introgressive hybridization? *BMC Evolutionary Biology*, *9*(1), 83. https://doi.org/10.1186/1471-2148-9-83

Zinner, D., Keller, C., Nyahongo, J. W., Butynski, T. M., de Jong, Y. A., Pozzi, L., . . . Roos, C. (2015). Distribution of mitochondrial clades and morphotypes of baboons Papio spp. (Primates: Cercopithecidae) in Eastern Africa. *Journal of East African Natural History*, *104*(1–2), 143–168. https://doi.org/10.2982/028.104.0111

Authors and Affiliations

Maria Joana Ferreira da Silva[1,2,3] · Paula Tralma[4] · Ivo Colmonero-Costeira[1,2,3,5] · Mariana Cravo-Mota[1,2] · Rassina Farassi[6,7,8,14] · Philippa Hammond[8] · Lynn Lewis-Bevan[8,9] · Marion K. Bamford[10] · Dora Biro[11] · Tina Lüdecke[8,12] · Jacinto Mathe[8] · René Bobe[8,13,14] · Cristian Capelli[15] · Susana Carvalho[1,2,8,13,14] · Felipe I. Martínez[16]

✉ Maria Joana Ferreira da Silva
ferreiradasilvamj@cardiff.ac.uk

Paula Tralma
paula.tralma.p@gmail.com

Ivo Colmonero-Costeira
ColmoneroCosteiraI@cardiff.ac.uk

Mariana Cravo-Mota
marianacpm@gmail.com

Rassina Farassi
rassinafarassi@gmail.com

Philippa Hammond
philippajhammond@gmail.com

Lynn Lewis-Bevan
lynn.lewis-bevan@biology.ox.ac.uk

Marion K. Bamford
Marion.Bamford@wits.ac.za

Dora Biro
dbiro@ur.rochester.edu

Tina Lüdecke
Tina.Luedecke@mpic.de

Jacinto Mathe
jacintomathe@gmail.com

René Bobe
rene.bobe@gorongosa.net

Cristian Capelli
cristian.capelli@unipr.it

Susana Carvalho
susana.carvalho@gorongosa.net

Felipe I. Martínez
fmartinezl@uc.cl

[1] CIBIO, Centro de Investigação Em Biodiversidade E Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Campus de Vairão, 4485 - 661 Vairão, Portugal

[2] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485 - 661 Vairão, Portugal

[3] ONE Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, The Museum Ave, Cardiff CF10 3 AX, UK

[4] Departamento de Arqueología, Universidad Austral de Chile, Los Ríos, Chile

[5] CIAS, Research Centre for Anthropology and Health, Department of Life Sciences, University of Coimbra, Coimbra, Portugal

[6] Departamento de Ciências da Vida, Centre for Functional Ecology, Universidade de Coimbra, Coimbra, Portugal

[7] Gorongosa National Park, Sofala, Mozambique

[8] Primate Models for Behavioural Evolution Lab, Institute of Human Sciences, School of Anthropology and Museum Ethnography, University of Oxford, Oxford, UK

[9] Department of Biology, University of Oxford, Oxford, UK

[10] Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

[11] Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

[12] Emmy Noether Group for Hominin Meat Consumption, Max Planck Institute for Chemistry, 55128 Mainz, Germany

[13] Department of Science, Gorongosa National Park, Sofala, Mozambique

[14] ICArEHB, Interdisciplinary Centre for Archaeology and the Evolution of Human Behaviour, Universidade do Algarve, Faro, Portugal

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

[16] Escuela de Antropología, Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile, Santiago, Chile