Extensive population structure highlights an apparent paradox of stasis in the impala (*Aepyceros melampus*)

Running title: **Population structure and stasis in impala**

Genís Garcia-Erill[1,2,#] , Xi Wang[1], Malthe S. Rasmussen[1], Liam Quinn[1], Anubhab Khan[1], Laura D. Bertola[1], Cindy G. Santander[1], Renzo F. Balboa[1], Joseph O. Ogutu[3], Patrícia Pečnerová[1], Kristian Hanghøj[1], Josiah Kuja[1], Casia Nursyifa[1], Charles Masembe[4], Vincent Muwanika[5], Faysal Bibi[6], Ida Moltke[1], Hans R. Siegismund[1], Anders Albrechtsen[1#], Rasmus Heller[1#]

[1] Department of Biology, University of Copenhagen, Copenhagen, Denmark

[2] Bioinformatics Research Center, Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

[3] Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

[4] College of Natural Sciences, Makerere University, P. O. Box 7062, Kampala, Uganda

[5] College of Agricultural and Environmental Sciences, Makerere University, P. O. Box 7062, Kampala, Uganda

[6] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, 10115 Berlin, Germany

[#] correspondence to: rheller@bio.ku.dk, aalbrechtsen@bio.ku.dk, genis.grcerll@gmail.com

Abstract

Impalas are unusual among bovids because they have remained morphologically similar over millions of years – a phenomenon referred to as evolutionary stasis. Here, we sequenced 119 whole genomes from the two extant subspecies of impala, the common (*Aepyceros melampus melampus*) and black-faced (*A. m. petersi*) impala. We investigated the evolutionary forces working within the species to explore how they might be associated with its evolutionary stasis as a taxon. Despite being one of the most abundant bovid species, we found low genetic diversity overall, and a phylogeographic signal of spatial expansion from southern to eastern Africa. Contrary to expectations under a scenario of evolutionary stasis, we found pronounced genetic structure between and within the two subspecies with indications of ancient, but not recent, gene flow. Black-faced impala and eastern African common impala populations had more runs of homozygosity than common impala in southern Africa, and, using a proxy for genetic load, we found that natural selection is working less efficiently in these populations compared to the southern African populations. Together with the fossil record, our results are consistent with a fixed-optimum model of evolutionary stasis, in which impalas in the southern African core of the range are able to stay near their evolutionary fitness optimum as a generalist ecotone species, whereas eastern African impalas may struggle to do so due to the effects of genetic drift and reduced adaptation to the local habitat, leading to recurrent local extinction in eastern Africa and re-colonization from the South.

Keywords: evolutionary stasis, phylogeography, population genetics, impala, whole genome sequencing

Introduction

The impala (*Aepyceros melampus*) belongs to the species-rich Bovidae (Artiodactyla). It has a unique combination of morphological traits that rendered its phylogenetic position challenging until the advent of phylogenomics, which consistently placed it as a basal branch

among the antilopine bovids (Antilopinae), together with the genera *Nesotragus* and *Neotragus* (Chen et al., 2019; Hassanin et al., 2012). The impala is the sole extant member of a >10 million-year-old bovid lineage (Chen et al., 2019), and has traditionally been placed in its own tribe, Aepycerotini (Groves & Grubb, 2011). The impala is an example of evolutionary stasis (Eldredge et al., 2005; Vrba, 1984) in a family of otherwise highly radiating taxa, by virtue of its low rates of morphological and taxonomic diversity in the fossil record (Gentry, 1985; Vrba, 1984). The oldest fossils assigned to *Aepyceros* are ≈7 million years old (Geraads, 2019), and specimens assigned to *Aepyceros shungurae*, which appeared in the fossil record between 4 and 3 million years ago, are difficult to distinguish morphologically from modern impalas based on the available cranial material (Gentry, 1985; Vrba, 1984). Such evolutionary stasis is unusual among bovids, making the impala an interesting lineage in which to investigate evolutionary dynamics and to assess possible links between short-term evolutionary dynamics and long-term evolutionary stasis during a time when many other African bovid lineages diversified rapidly.

Evolutionary stasis is characterised as little to no significant morphological changes in a lineage across millions of years (Eldredge et al. 2005). The concept originates from palaeontology, especially in the context of the hypothesis of punctuated equilibrium, which argues that the evolution of most species takes place during rapid bouts of change, which interrupt long periods of stasis (Eldredge & Gould, 1972; Gould & Eldredge, 1977). To our knowledge there are no studies that link population genetic processes with deep-time evolutionary stasis, therefore it is relatively poorly understood which specific combination of population genetic processes characterises evolutionarily stable species, such as the impala. However, theoretical papers have highlighted the roles of stabilizing selection and gene flow in maintaining stasis (Eldredge et al., 2005). First, stabilizing selection tends to maintain populations close to their fitness optimum, a process that is more efficient in larger populations than in smaller ones. If the fitness optimum remains relatively stable, this type of selection – provided it can overcome genetic drift – can lead to evolutionary stasis in the population (S. Estes & Arnold, 2007). Second, even if some local populations deviate from the species-wide fitness optimum or begin adapting to local conditions, gene flow tends to homogenize the species as a whole (Hunt & Rabosky, 2014). Collectively, these factors suggest two testable microevolutionary conditions for evolutionary stasis: (i) stasis is more likely when

population structure is limited and populations are not isolated for extended periods, and (ii) stasis is more likely in cases of relatively large effective population sizes (Voje, 2018).

The impala, an ecological generalist, inhabits woodland-grassland ecotones and requires access to water. They exhibit a flexible dietary composition, being able to browse and graze in varying proportions, depending on the season and availability of resources (Fritz & Bourgarel, 2013). Vrba (Vrba, 1980, 1984) posited that the generalist ecology of the impala has been crucial to its morphological stability over extended evolutionary periods. This flexibility enables the impala to cope with variable environments rather than specifically adapt to them. This contrasts with, e.g., the highly variable morphologies found in the bovid specialist tribe Alcelaphini, including species such as the wildebeests and hartebeests. Impalas are usually classified into two subspecies, the common impala (*A. melampus melampus*) and the black-faced impala (*A. melampus petersi*) (Fritz & Bourgarel, 2013). Together, they are distributed throughout the savannah bushlands of southern and eastern Africa, with the common impala having by far the largest distribution and the black-faced impala being restricted to northern Namibia (Figure 1). The species has a combined estimated population size of 2 million individuals, and is the most abundant ungulate in parts of southern Africa, with the black-faced impala amounting to just ≈3300 (perhaps 4300) (Fritz & Bourgarel, 2013). Previous studies using mtDNA concluded that the main genetic structure within the common impala is characterised by a separation of southern from eastern populations, and that eastern populations appeared to have a younger origin (Lorenzen et al., 2006; Nersting & Arctander, 2001). Microsatellite-based studies initially did not find any sign of hybridization between the common and the black-faced impala (Lorenzen & Siegismund, 2004), but a recent study identified several recent hybrids in Etosha National Park and nearby private game reserves (Miller et al., 2020).

To elucidate the evolutionary processes at work within impalas and to understand how these forces might have limited morphological changes from either i) originating and increasing in frequency locally in only one or a few populations, or ii) spreading across the species' range, we conducted a population genomic study. Specifically, we sequenced 119 whole genomes from both common and black-faced impalas across their range. To address (i), we inferred impala's genetic structure and admixture, inferred a phylogeographic model to explain the

122 distribution of genetic variation, and considered historical processes such as ancient
123 admixture and changes in population size. For (ii), we estimated historical population sizes
124 and genetic diversity, and compared the efficiency of natural selection in removing new non-
125 synonymous variants that arise in the different populations. We interpret these results in
126 conjunction with fossil record data, exploring various scenarios that could lead to
127 evolutionary stasis, and discuss the implications of our findings for impala conservation.

128 Results

129 We sequenced 119 impala samples from 13 different African localities, comprising 105
130 samples at low depth (2-4x) and 14 samples at medium depth (7-18x) of genome coverage.
131 Each sampling locality combines low-depth samples and a high-depth sample, which allows
132 the detection of any potential batch effect associated with differences in sequencing depth
133 (Lou & Therkildsen, 2022). Seven of these samples were from black-faced impala, including
134 one medium depth, while the remaining 112 were from common impala (Figure 1,
135 Supplementary Table 1). We mapped the sample reads to both an impala draft reference
136 genome (IMP, (Chen et al., 2019)) and a goat chromosome-level reference genome (ARS1,
137 (Bickhart et al., 2017)), and after strict filtering of sites retained 1.07 and 1.3 Gbp from each
138 respective genome for analyses (Supplementary Table S2). We used preferentially the data
139 mapped to the impala reference genome except for analysis sensitive to reference bias and
140 analyses that require information on chromosomal position of sites, where we used the data
141 mapped to the goat genome because it is an outgroup and it is a chromosome-level genome,
142 respectively (see Method section of each analysis for details). During sample quality control
143 we identified 1 low-depth sample with excessively high sequencing error rates
144 (Supplementary Figure S1) and 5 pairs of duplicate low-depth samples (Supplementary Figure
145 S2). Consequently, we excluded 6 samples from all downstream analyses, resulting in a final
146 dataset of 113 samples (Figure 1A).
147

148 Population structure, recent admixture and genetic diversity

149 A principal component analysis (PCA) separated the samples from the two subspecies on PC1
150 and differentiated the southern from the eastern African common impala on PC2 (Figure 1B).

151 Notably, common impalas from Selous were intermediate to southern and eastern Africa on
152 both PC1 and PC2. Higher PCs revealed further population structure within the southern and
153 eastern common impala localities (Supplementary Figure S3). Admixture analyses using
154 NGSadmix (Skotte et al., 2013) corroborated the presence of strong population structure,
155 requiring up to K=12 to obtain an overall good model fit using evalAdmix (Garcia-Erill &
156 Albrechtsen, 2020) (Figure 1C, Supplementary Figures S4 and S5). At K=12, samples from the
157 two Zambian localities, Kafue and Luangwa, showed elevated negative correlation of residuals
158 between them, suggesting they represent separate populations, despite being assigned to
159 the same cluster. Admixture at K > 12 did not converge after 100 independent runs, possibly
160 because the low sample sizes in the two localities prevented the model from assigning each
161 its own cluster at higher K. Finally, the captive common impala population in Ovita showed
162 some noise in individual residual correlations within the population, which could be due to
163 the presence of distantly related individuals and the recent founding of this population, which
164 consists of individuals originally from southern Africa (East, 1998). Otherwise the evalAdmix
165 results are consistent with all sampled localities having distinct and homogeneous genetic
166 ancestries.
167
168 Population pairwise F_{ST} values confirmed the existence of two broad groups consisting of
169 southern and eastern common impala populations, with Selous showing similar F_{ST} (≈0.2) to
170 both groups (Figure 1D). Within the southern populations, F_{ST} was relatively low, ranging from
171 0.04-0.10. In contrast, the eastern group exhibited a slightly wider range of F_{ST} values, from
172 0.04-0.24. This was reversed for D_{xy}, where southern populations showed higher D_{xy} (≈0.002)
173 between them than eastern populations (≈0.0011-0.0015). This indicates a more recent
174 divergence, but with greater drift between eastern populations, especially those at the
175 northern extreme of the species' range, such as in Lake Mburo and Samburu. F_{ST} for southern
176 versus eastern populations ranged between 0.18-0.33. Notably, the lower values were mainly
177 found between Selous or Ugalla and the southern populations. Moreover, the D_{xy} between
178 eastern and southern populations were of similar magnitude to the D_{xy} within southern
179 populations, again suggesting a recent divergence. Between black-faced and common impala
180 populations, F_{ST} ranged between 0.46-0.61, with the lowest values involving southern
181 populations, while D_{xy} was equally elevated with all common impala populations at ≈0.0033.
182

183 Genetic diversity, measured as per-individual heterozygosity, showed a clear geographical
184 trend after removing the low-depth samples with elevated sequencing error rates
185 (Supplementary Figures S1 and S6). The highest genetic diversity values are found in the
186 southernmost populations, Shangani and Chobe, and gradually decline northwards.
187 Surprisingly, despite being geographically restricted and less abundant, the black-faced
188 impala had genetic diversity on par with the southern common impala populations,
189 considerably higher than those in eastern Africa (Figure 2A). This pattern of gradually and
190 monotonically declining genetic diversity in common impala is highly suggestive of a
191 directional expansion (Henn et al., 2019). We explored this further by inferring a population
192 tree with TreeMix. When not allowing for migration, the tree shows a successive split of
193 eastern African common impala populations, roughly coinciding with their geographical
194 distance to southern Africa, with Selous showing the earliest split from the other eastern
195 populations (Figure 2B). This reinforces the notion that the eastern populations likely
196 originated from an expansion of the southern populations.
197
198 A geography-aware analysis of genetic barriers and connectivity, using EEMS (Estimating
199 Effective Migration Surfaces) on the non-captive populations, confirmed that some parts of
200 the impala range are characterised by barriers to gene flow (Figure 2C). One notable barrier
201 exists between black-faced impala and all other populations. Another separates the southern
202 and eastern common impala populations and a third strong barrier is located north of Selous
203 which possibly geographically overlaps with the Eastern Arc mountain chain. Additional, less
204 apparent barriers identified in the other population structure analyses included one
205 separating Kafue and Luangwa from Mana Pools and Chobe, overlapping with the Zambezi
206 river, and one separating Ugalla from Lake Mburo, overlapping with Lake Victoria. Hence,
207 some barriers correspond with major geographical features that likely serve as physical limits
208 to gene flow, whereas others – notably that between southern and eastern common impala
209 populations, are less clearly linked to present-day geographical features.
210

211 **Population divergence and ancestral admixture events**

212 We further used TreeMix to understand ancient gene flow between populations by inferring
213 population trees with admixture between branches. TreeMix has high residuals for up to five

214 admixture events, which suggests abundant gene flow events between populations, including
215 one involving the black-faced impala population from Etosha with its geographically closest
216 southern common impala in Chobe (Supplementary Figure S7). We used D-statistic analyses
217 to further explore the migration events, and used the f-branch statistic implemented in Dsuite
218 (Malinsky et al., 2021) as a visual summary of all combinations of D-statistics with topologies
219 compatible with the TreeMix tree without migration (Figure 2B). This revealed further
220 evidence of multiple admixture events (Figure 3A), including between Selous and Ugalla and
221 the southern populations, and between the two Zambian populations and eastern
222 populations. Moreover, the f-branch statistic also showed evidence of admixture between
223 the common impala in Chobe and the black-faced impala. To further understand this
224 observation, we visualised the D-statistic analyses of the type (((P1,P2),black-faced),Goat),
225 with all possible pairs of common impala populations as P1 and P2. These confirmed that
226 Chobe shares more alleles with the black-faced impala than other common impala
227 populations do, and that southern populations generally share more alleles with black-faced
228 impala than eastern ones (Supplementary Figure S8). Consistent with a Chobe to black-faced
229 impala introgression, a mtDNA tree showed that two black-faced impala individuals carried
230 mtDNA with common impala haplotypes, with Chobe individuals being their closest
231 haplotypes in both cases (Supplementary Figure S9). We used *qpgraph* to quantify the
232 amount of gene flow. To focus on the most important features of the species history and
233 avoid fitting an overly complex model, we chose to include only the following populations:
234 Etosha (black-faced), Chobe (southern common), Shangani (southern common) and Masai
235 Mara (eastern common). We estimated that 16% of black-faced impala ancestry introgressed
236 from Chobe into black-faced impala (Figure 3B). While the admixture graph with Chobe to
237 black-faced introgression fitted the observed f-statistics without any significant deviation, 3
238 other fundamentally different admixture graphs with a single migration event were equally
239 well supported by the allele sharing data (Supplementary Figure S10). Our choice of the most
240 likely admixture graph out of the 4 candidates was guided by its ability to best explain the
241 presence of common impala mtDNA haplotypes in black-faced impala individuals.

242 Demographic history

243 We used PSMC to infer the effective population sizes through time from genotype calls of all
244 high and medium-depth samples. To explore the impact of sequencing depth and optimal

245 filtering settings, we also included a previously published high-depth sample from the
246 ruminant genome project (RGP) that was sampled in Laikipia, Kenya (Supplementary Figure
247 S11). PSMC analyses showed a declining trend in population sizes since about 1 million years
248 ago, and highly similar population histories across all impala populations until around 50 kya
249 (Figure 4A). Around 50 kya the southern common impala populations started to increase in
250 size, whereas the eastern as well as the black-faced population continued to decline. We
251 caution that population structure and gene flow can confound population size history
252 inference (Li & Durbin, 2011). Even so, a major split between the demographic histories seems
253 to have occurred around 50 kya. To further corroborate this finding and date the main
254 divergence events in the impala, we carried out demographic history modelling using
255 fastsimcoal2 (Excoffier et al., 2021) with four of the same populations used in the *qpgraph*
256 analysis, excluding the population from Tsavo, Kenya, because it is expected to have a very
257 similar demographic history to the population from Masai Mara and would thus add
258 unnecessary complexity to the model. We tested four different models (Supplementary
259 Figure S12 and Supplementary Tables S2, S4, S5 and S6), and present results for the two with
260 the highest maximum likelihoods. The split time between southern and eastern common
261 impala was estimated at 56 kya, with gene flow occurring at 5 kya, assuming the gene flow
262 was a single admixture pulse (Figure 4B). The divergence time between black-faced and
263 common impala was estimated at 59 kya, with 14% gene flow from Chobe to Etosha occurring
264 around 2 kya. Additionally, we detected highly asymmetric gene flow from Masai Mara to
265 the southern population ancestral to Shangani and Chobe (Figure 4B) and obtained similar
266 results between the southern ancestral population and Masai Mara under continuous gene
267 flow assumptions (Figure 4C). To account for uncertainties in impala's generation time, we
268 also scaled the demographic history timing with an alternative estimate of generation time.
269 The alternative generation time of 4.03 compared to the would result in younger dating of
270 the different events in impala's demography, being reduced by a factor of ≈0.71
271 (Supplementary Figure S13).

272 Recent population size and strength of selection

273 To investigate more recent demographic events and their potential effects on genomic
274 diversity, we first inferred the lengths and total proportion of runs of homozygosity (ROHs) in
275 the different individuals. We based these analyses on imputed genotype calls mapped to the

276 goat reference genome, after confirming with a PCA that the imputed data preserved the
277 same main data features as those estimated from genotype likelihoods (Supplementary
278 Figure S14 and Fig 1B). We found that between 1-30% of the genome resides in ROHs across
279 individuals, displaying a clear geographic trend. Specifically, the eastern common impala
280 populations, particularly those in Samburu and Lake Mburo, along with black-faced impala,
281 exhibit progressively more ROHs than the southern common impala populations (Figure 5A).
282 We also found several individuals with a large proportion of their genome in long ROHs (>10
283 Mbp) in the black-faced impala population from Etosha, the captive common impala
284 population in Ovita and the common impala population in Luangwa, suggesting cases of
285 recent inbreeding within these populations. Linkage disequilibrium (LD) decay curves for each
286 population corroborated this trend, revealing a slower LD decay in the eastern common
287 impala populations, especially in Lake Mburo (Figure 5B), indicative of a reduced effective
288 population size in recent times. The reduced effective population size of the black-faced
289 impala and the eastern common impala can impair the effectiveness of selection. To explore
290 the efficiency of natural selection in removing new functional mutations, which are
291 predominantly harmful, we estimated the ratio of derived alleles at protein coding sites.
292 Specifically, we compared the number of derived alleles at 0-fold degenerate sites, where any
293 substitution alters the encoded amino acid, to those at 4-fold degenerate sites, where
294 substitution preserves the amino acid, across each population. This ratio (Supplementary
295 Figure S15) serves as a proxy for the effectiveness of natural selection in each population
296 (Clemente & Vogl, 2012). This ratio was lower in Selous than in the southern common impala
297 populations, and both were markedly lower than those for the eastern common impala
298 populations and the black-faced impala (Figure 5C). We observed a correlation of the ratio
299 between the total number of derived alleles in the population with the preceding ratio
300 (Supplementary Figure 16), which suggests error rates might bias the ratio by having a higher
301 relative impact on the more evolutionary constrained and less diverse 0-fold degenerate sites.
302 This led us to exclude the population from Shangani, Zimbabwe from the plot, and we warn
303 it might potentially impact the results for the eastern population from Selous, Tanzania, and
304 the black-faced population from Etosha, Namibia.

Discussion

A synthesis of the evolutionary processes within the impala

We show that impalas have extensive population genetic structure and can be distinctly classified into three major lineages, each with substantial genetic differentiation: the black-faced impala, the southern common impala and the eastern common impala. Selous, the southernmost population among the eastern common impalas, emerges as a basal branch of this group. It occupies an intermediate position in the PCA plot and has a relatively high F_{ST} (≈0.2) with both the southern and the other eastern populations, indicating it has a unique history and may represent a distinct phylogeographic lineage.

While the estimated divergence between the southern and eastern common impala populations is temporally close to that between black-faced and common impala (≈50-55 kya and ≈52-58 kya, respectively), our models also indicate connectivity post divergence. This suggests a weakly structured population rather than an instantaneous split. The onset of the southern-eastern divergence may have occurred as common impalas began to colonise Eastern Africa from a southern Africa refugium (Lorenzen et al. 2006). This scenario, considering the inherent uncertainties of demographic modelling, accords with the suggestion that common impalas were certainly present in East Africa around 57 kya, and that they may have replaced a morphologically distinct, now-extinct *Aepyceros* population in eastern Africa around this time (Faith et al., 2014). In more recent times, the clear structure and lack of recent admixture between populations imply minimal ongoing gene flow between populations.

The genetic diversity in impalas is exceptionally low (range 0.001-0.002), relative to other African mammals such as the African buffalo (0.0026-0.0051, (Quinn et al., 2023), warthog (0.0025-0.0035, (Garcia-Erill et al., 2022), African leopard (0.0020, (Pečnerová et al., 2021), and waterbuck (0.0030-0.0055, (Wang et al., 2022). This is surprising, given that impalas are both very abundant (East, 1998) and evolutionarily successful, traits typically associated with high population sizes and genetic diversity. A PSMC analysis revealed that impalas have had moderate to low effective population sizes for most of the Pleistocene, which might offer

335 some explanation. One possible explanation is the species' highly polygynous social structure,
336 characterised by a single territorial male monopolising mating with harems of up to 50-100
337 females (R. Estes, 1991, 1993). This extreme skew in effective sex ratios could significantly
338 reduce the effective population size. Further research is required to ascertain if this factor
339 alone can explain the surprisingly low genetic diversity in the species. Alternatively, historical
340 confinement to a smaller southern refugium could explain impala's historically low
341 population sizes.
342
343 Despite the evident strong structure and limited recent gene flow, we found signals of
344 extensive ancient admixture across all impala lineages. This includes notable introgression
345 from the common impala to the black-faced impala, with the latter's ancestry comprising 14-
346 16% common impala genetic ancestry. We inferred this introgression to have occurred
347 around 2 kya, which would explain why it could not be detected using methods aimed at
348 detecting recent admixture on some of the same samples (Lorenzen & Siegismund, 2004).
349 Consistent with this, we found two black-faced individuals carrying common impala mtDNA
350 haplotypes, as was also found in an earlier study (Nersting & Arctander, 2001). We therefore
351 conclude that introgression occurred naturally under climatic conditions that brought the two
352 subspecies into contact, rather than as a result of recent translocations of common impala to
353 Namibia, which has been a concern for the genetic integrity of the subspecies (Lorenzen &
354 Siegismund, 2004; Miller et al., 2020). This introgression might have contributed to the
355 relatively high genetic diversity found in the black-faced impala, despite its recent low
356 population size. However, the Etosha population of the black-faced impala consisted of a
357 small group; approximately 60 individuals, before 1970 (Joubert, 1971). The population was
358 supplemented by 180 individuals from three translocations during 1968-1971 from different
359 areas of its range and three more translocations of a total of 194 from 1970 to 1979 (Green
360 & Rothstein, 1998; Joubert, 1971).

361 **Are microevolutionary patterns reconcilable with evolutionary stasis in impala?**

362 The pronounced genetic structure and low genetic diversity observed in impalas suggest that
363 they do not meet the two ideal microevolutionary conditions (panmixia and large effective
364 population sizes) for evolutionary stasis. Hence, impalas would seem to exemplify the so-
365 called 'paradox of stasis', which describes the discrepancy between the fossil-based inference

366 of morphological stasis and the potential for rapid genetic change observed in many species
367 (Eldredge et al., 2005). In what follows, we integrate our findings on impala population
368 genetics and phylogeography with their fossil record in an attempt to explore and potentially
369 resolve this apparent paradox of stasis in the species.
370
371 Evolutionary stasis in impalas has been attributed to impalas being "superbly adapted to their
372 habitat" (Kingdon, 2013). This hypothesis suggests that impalas have reached a fixed-
373 optimum of fitness, where stabilizing selection maintains this peak, and where the fitness
374 optimum remains relatively stable over time, consistent with the notion of prevalent stable
375 adaptive zones suggested by Estes & Arnold (2007). Consequently, most new phenotype-
376 altering mutations – especially those related to these adaptations – would face negative
377 selection. This selection process should be most efficient in large populations and less so in
378 smaller, isolated ones, due to the influence of genetic drift (Hunt & Rabosky, 2014; Voje,
379 2018). Our results suggest that, indeed, in the eastern common and black-faced impala
380 populations, selection was less efficient in removing functional mutations than in the
381 southern common impala populations. These observations are consistent with the inferred
382 phylogeography indicating that eastern common impala populations descend from a
383 relatively recent spatial expansion from southern Africa, and therefore have experienced
384 more genetic drift. Interestingly, this is also in agreement with morphological variation within
385 the impala (Bastos-Silveira & Lister, 2007), which was subsequently interpreted as evidence
386 that the extant eastern African impala populations originated from a subset of the southern
387 African populations (Reynolds, 2010), and with fossil-based findings suggesting that impalas
388 in eastern Africa were subject to more variable conditions, forcing them to evolve new
389 adaptations or go extinct (Faith et al., 2014; Reynolds, 2007). From this combined evidence,
390 it appears the evolutionary stasis in impalas is consistent with a fixed-optimum hypothesis
391 where southern Africa comprises the core region for impalas. Here, comparatively high
392 effective population sizes (but see above) have allowed populations to stay near a stable
393 fitness peak as an ecotone generalist herbivore under the influence of stabilizing selection. In
394 contrast, eastern African impalas have had more volatile demographic histories, greater
395 morphological variability (Reynolds, 2010) due to less efficient stabilizing selection, and have
396 possibly gone locally extinct at least once (Faith et al., 2014). Our results indicate that, at least
397 during the recent geological past, the extant impala's optimal habitat was to be found in

southern Africa, where it persisted and from where it recolonized eastern Africa, perhaps more than once, after local extirpation events there. Why remains unknown, and climate change, competition with other herbivores, or even hominin-related disturbances are all possible factors worth future investigation. Similarly, we cannot determine for how long southern Africa has acted as a refugium for the impala lineage, nor whether the optimum shifted geographically during the millions of years of its existence. The existence of a Pleistocene southern African refugium for savannah-adapted ungulates with present-day southern and eastern African distributions is supported by phylogeographic evidence from several other species (Lorenzen et al., 2012), including blue wildebeest (Arctander et al., 1999), plains zebra (Pedersen et al., 2018) and common eland (Lorenzen et al., 2010).

Hence, we outline a scenario that could offer a resolution to the paradox of stasis in impalas. The inferred phylogeography, combined with evidence that selection is less efficient in eastern Africa, could explain why impalas manage to stay at a fitness optimum in southern Africa, but fail to do so in eastern Africa. Populations that deviate phenotypically from the fitness optimum may experience higher turnover and local extinction risks (Eldredge et al., 2005), and this may be exacerbated if such “pioneer” populations tend to inhabit areas that are more environmentally volatile. This would cause phenotypically divergent populations to be more temporally transient, and therefore appear infrequently or not at all in the fossil record, leading to an appearance of macroevolutionary stasis despite dynamic microevolutionary changes. Hence, this tentative explanation reconciles a number of key observations in impalas: dynamic microevolutionary change, differential selection across the species range, the existence of a southern African refugium, and the fossil record of the impala. However, this explanation remains a hypothetical reconciliation of the apparent paradox of stasis in impala. It assumes, e.g., that the inferred microevolutionary dynamics – though only informative for the last 100,000 years or so – are representative of evolutionary events on longer time scales, which cannot be taken for granted (Olson, 2024).

Conservation implications

Our demographic analyses suggest that certain impala populations, in particular the eastern populations and black-faced impala, have maintained small sizes for tens to hundreds of thousands of years. Furthermore, the black-faced population in Etosha and the eastern

common impala from Samburu and Lake Mburo show signs of recent population declines, as evidenced by an increase in the proportion of the genome covered by relatively large ROHs (>2 Mb) and in the extent of linkage disequilibrium. In addition, the eastern and black-faced populations have a markedly higher accumulation of functional mutations, indicating less efficient selection. This was especially true for the black-faced population from Etosha, and Samburu common impala, the northernmost population in Africa. Given these findings, it is crucial to monitor these populations at either geographical extreme for potential inbreeding depression and increased genetic load.

The primary conservation concern for the black-faced impala has been the threat of genetic swamping by the much more widespread and abundant common impala. This threat was exacerbated by the translocation of common impala to game farms near Etosha National Park, where the black-faced impala also occurs (Green & Rothstein, 1998). While we found no evidence of recent gene flow between the two subspecies, we identified a relatively large historical introgression event from southern common impala into the black-faced impala, ≈5 kya. This event may have contributed to the black-faced impala's surprisingly high genetic diversity, despite their historically much lower population sizes than the common impala. Our finding of natural introgression between the two subspecies under certain conditions adds complexity to the task of defining discrete evolutionary units for conservation.

Materials and methods

Sample Collection and Laboratory Protocol

We used 119 samples of impala tissue, mainly dried skin, collected during 1994–1999 in Uganda, Kenya, Tanzania, Zambia, Zimbabwe, Botswana and Namibia. All samples, except those from Ovita Game Ranch, were analysed using other genetic markers in previous studies (Lorenzen et al., 2006; Lorenzen & Siegismund, 2004; Nersting & Arctander, 2001) where sample collection and storage is described. The Ovita Ranch population has not previously been analysed, but samples were collected as part of the same project and stored in the same way as the remaining samples.

457 We used the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA, USA) for DNA
458 extraction following the manufacturer's protocol. We then added RNase to the samples to
459 remove all RNA from the samples. Finally we measured DNA concentrations with Qubit 2.0
460 Fluorometer (ThermoFisher Scientific) using a dsDNA BR Assay Kit and a Nanodrop One
461 spectrophotometer (ThermoFisher Scientific), and used gel electrophoresis to check the
462 quality of the genomic DNA.

463 Sequencing

464 Library preparation and sequencing was performed at BGI, Shenzhen, China. We used Illumina
465 150 bp paired-end sequencing. Fourteen samples each representing different sampling
466 locations were sequenced at medium depth (7-18X) using the HiSeq2500 platform while the
467 rest of the samples were sequenced to 2-4X using the NovaSeq platform (Supplementary
468 Table S1).
469
470 We also included a high depth sample that was previously published as part of the Ruminant
471 Genome Project (RGP, (Chen et al., 2019)), and that had been used to assemble the impala
472 draft reference genome (BioSample ID SAMN08714480). Based on an initial fastQC screening
473 we selected 5 SRA runs that showed best scores on base composition, quality scores and GC
474 percentage. All runs were short read paired end data sequenced with HiSeq X Ten and with
475 insert sizes of 500 bp (SRR6860888, SRR6860889 and SRR6860891), 600 bp (SRR6860892) and
476 900 bp (SRR6860896). This data was only used for inference of demographic history with the
477 pairwise sequentially Markovian coalescence (PSMC, see below).
478
479 Initial QC was performed on the raw reads using FastQC (Andrews, 2010) and MultiQC (Ewels
480 et al., 2016).
481

482 Mapping

483 First we used NGSmerge (Gaspar, 2018) on the raw read sequence data to merge all paired
484 reads that overlapped by 11 bp or more. Following this, we separately mapped the merged
485 reads as single-end reads and the non-merged paired-end reads with bwa mem v0.7.17 (Li,

2013) in single-end and paired-end mode, respectively. We then used samtools v. 1.9 to mark and remove duplicate reads, removed low quality alignments (flag -F 3852) and non-merged reads that were not properly mapped (flag -f 3). Finally, we combined both non-merged and merged reads into a single bam file for each sample. We applied this data processing and mapping pipeline using two different reference genomes: the draft impala reference genome IMP (accession GCA_006408695.1, (Chen et al., 2019)) and the goat reference genome ARS1 (refSeq accession GCF_001704415.2, (Bickhart et al., 2017)).

Common filter settings

For all analyses, we excluded bases with base call quality below 30 and reads with mapping quality below 30. For all analyses that involved estimation of genotype likelihoods with ANGSD (Korneliussen et al., 2014), we used the GATK genotype likelihood model (-gl 2; (McKenna et al., 2010)).

Reference genome masks

We masked regions in each of the reference genomes by removing data located in regions that are most likely to contain mapping and genotyping errors (Pečnerová et al., 2021). These were identified with different approaches described below. All approaches were applied with the same settings for both the reference genomes, unless explicitly stated otherwise.

Scaffold length

For the impala reference genome, after an initial check of the cumulative amount of sequence contained in scaffolds below a certain length, we excluded all scaffolds shorter than 100 kb, which retained around 92% of all sequences in the reference genome (Supplementary Figure S17). For the goat genome, we discarded all scaffolds not assembled into chromosomes.

Repeats

For both reference genomes, we removed all regions annotated as repeats using the corresponding publicly available repeat annotations (Bickhart et al., 2017; Chen et al., 2019).

511 ## Mappability

512 We estimated mappability across the genome and excluded all sites with mappability lower
513 than 1. We used genmap (Pockrandt et al., 2020) to infer mappability, using a read length of
514 100 (-K 100) and a maximum number of mismatches of 2 (-E 2).

515 ## Sex chromosomes

516 For most analyses we kept only autosomal chromosomes. For the impala genome, we used a
517 preliminary version of SATC (Nursyifa et al., 2022) to jointly identify sex associated scaffolds
518 and sample sex based on differences in normalised depth between the sexes, and excluded
519 all scaffolds with a significant difference in normalised depth between the two sexes.

520 ## Depth

521 We excluded sites with excessively low or high depth. We estimated local depths using ANGSD
522 for two different groupings of samples, one pooling together all low depth samples, excluding
523 only sample 2145 with excessive error rates (Supplementary Figure S1, see methods below),
524 and another pooling together all medium-high depth samples. We then removed from
525 analyses any site with combined depth below 0.5 × median depth or above 1.5 × median
526 depth for either of the two groups.

527 ## Excess heterozygosity

528 We estimated a preliminary dataset of genotype likelihoods for the common impala samples
529 after removing samples with high error rates and 5 duplicate samples (see below). We
530 estimated genotype likelihoods with ANGSD, calling variable sites as those with a p-value
531 lower than 1e-6 (-snp_pval 1e-6) and with a minor allele frequency (MAF) higher than 0.05 (-
532 -maf 0.05). We used these as input for PCAngsd (Meisner & Albrechtsen, 2018) to conduct a
533 Hardy-Weinberg equilibrium (HWE) test (Meisner & Albrechtsen, 2019), which estimates per
534 site inbreeding coefficients *F* in the range of 1 (total excess of homozygosity) and -1 (total
535 excess of heterozygosity), while using individual frequencies estimated from the first principal
536 components to account for population structure. We used 6 principal components to control
537 for population structure, and only included common impalas because the test cannot account
538 for population structure between highly diverged populations, and thus would not be

539 guaranteed to work well when combining common and black-faced impalas. We identified
540 sites with excess heterozygosity as those showing an F value lower than -0.9 and a p-value for
541 the HWE test lower than 1e-6, and excluded all sites within 10 kbp from them.

542 Sample quality control

543 Error rates estimation

544 We estimated error rates using the 'perfect sample' approach implemented in ANGSD
545 (Korneliussen et al., 2014). We used the data mapped to the goat to first generate a consensus
546 fasta sequence for a medium depth common impala sample from Tsavo, Kenya (sample ID
547 54) and for a medium depth black-faced impala from Etosha, Namibia (sample ID 2076),
548 filtering reads with mapping quality below 30 and bases with base quality below 30. We used
549 each of these two as the 'error free' individuals to calculate error rates for all samples, with
550 the goat reference genome as the outgroup. Under the assumption that the perfect individual
551 and the target sample have had equal mutation rates, the method estimates error rates
552 (which are compounded of both sequencing error rates but also errors due to DNA damage)
553 as an excess of mismatches of the sample to the outgroup relative to the mismatches of the
554 perfect sample (Orlando et al., 2013). We initially did the analyses using the goat autosomal
555 chromosomes and no further filtering of sites, using only sample 54 as 'perfect'. We
556 subsequently repeated the error rate estimation using all site filters and both sample 54 and
557 sample 2076, after observing a systematic reduction in error rates for black-faced impala
558 samples when using the common impala as 'perfect individual' (Supplementary Figure S1).

559 Duplicate and relatedness estimation

560 We used an allele-frequency method (Waples et al., 2019) that estimates three statistics
561 reflecting relatedness between pairs of samples from their 2-dimensional site frequency
562 spectrum (2DSFS), to estimate relatedness between all pairs of samples within each sampling
563 locality. We used ANGSD to estimate site allele frequency (SAF) likelihoods for each individual
564 mapped to the impala reference genome, from which we estimated the 2DSFS between each
565 pair using realSFS (Nielsen et al., 2012).

566 # Genotype likelihoods

567 We estimated genotype likelihoods for all 113 impala samples that passed sample filtering,
568 from the data mapped to the impala reference genome after applying all site filters. We called
569 SNPs with ANGSD (Korneliussen et al., 2014) using a SNP p-value of 1e-6 (-snp_pval 1e-6) and
570 a minor allele frequency of 0.05 (-maf 0.05), inferring major and minor alleles from the
571 genotype likelihoods (-doMajorMinor 1).

572 # Genotype calling

573 Based on the data mapped to the goat reference genome, we called genotypes for the
574 medium and high depth samples using bcftools 1.13 (Danecek et al., 2021) filtering to
575 minimum base and mapping quality 30. Based on the raw calls, we applied the site filters and
576 removed multiallelic sites as well as all indels. In addition, we masked all genotypes based on
577 less than six reads as well as any heterozygous calls for which either of the two alleles were
578 supported by less than three alternate alleles.

579 # Imputation

580 We called genotypes on all 113 samples kept after sample filtering, using imputation to refine
581 the genotype calls. We first used bcftools v1.10 (Danecek et al., 2021) to call genotypes and
582 genotype likelihoods on the data mapped to the goat genome, performing joint calling on all
583 samples with bcftools' multiallelic caller, keeping only variable sites, and subsequently filtered
584 to remove any site within 10 or less bp of an indel and keep only diallelic SNPs. We then
585 extracted the genotype likelihoods for common SNPs, using ANGSD to estimate allele
586 frequencies and filter sites with maf lower than 0.05, and used the genotype likelihoods as
587 input for imputing genotypes with BEAGLE3 (Browning & Browning, 2009). We subsequently
588 kept only sites with an imputation score (estimated r^2) higher than 0.9, and called the
589 genotype with the maximum posterior probability for each site and individuals, setting to
590 missing sites where the maximum posterior probability was lower than 0.9.

591 # SFS estimation

592 We estimated SAF files for all populations using all samples passing sample filtering within
593 each of the populations, using the data mapped to the impala reference genome and using
594 the allele in the reference genome to define the ancestral states, and otherwise using the
595 above described common settings. Using the SAF files, we ran winsfs (Rasmussen et al., 2022)
596 with default parameters to estimate the frequency spectra for all population pairs.

597 # Inference of population structure

598 ## Principal component analysis

599 We used PCAngsd (Meisner & Albrechtsen, 2018) to estimate the covariance matrix between
600 all 113 samples from the common SNP genotype likelihood dataset described above. We used
601 7 principal components for estimating individual frequencies in PCAngsd's iterative algorithm
602 to estimate the sample covariance matrix, after testing different numbers of principal
603 components (7, 10 and 12) and visually assessing the principal components obtained. As the
604 eigenvectors of the estimated covariance matrix were not sensitive to variation above 7, we
605 used this value for the analyses presented here.

606 ## Estimation and evaluation of admixture proportions

607 We used NGSadmix (Skotte et al., 2013) to estimate admixture proportions from the common
608 SNP genotype likelihood dataset mapped to the impala reference genome, assuming from
609 K=2 to K=13. For each K, we ran independent optimization runs until either the results
610 converged, defined as having the 3 runs with the highest likelihood within 3 or less log
611 likelihood units of each other, or 100 independent runs were reached without convergence,
612 which was the case when assuming K=13. For the results that converged, we evaluated the
613 model fit of the maximum likelihood run for each K by estimating the individual pairwise
614 correlation of residuals with evalAdmix (Garcia-Erill & Albrechtsen, 2020).

615 # Estimation of D_{xy} and F_{ST}

616 Using the population SAF files estimated as previously described, based on the data mapped
617 to the impala genome, we used realSFS to estimate a 2DSFS between each population, that
618 we subsequently used as input to estimate pairwise D_{xy} and F_{ST} using Hudson's estimator
619 (Bhatia et al., 2013) between each population pair with the *sfs* CLI tool (Rasmussen et al.,
620 2024) available at https://github.com/malthesr/sfs.

621 # Estimation of heterozygosity

622 We estimated per sample heterozygosity by first estimating individual SAF files with ANGSD
623 from the data mapped to the impala reference genome, with the same settings as previously
624 described. We then used realSFS to estimate the sample SFS, and estimated heterozygosity
625 as the inferred number of heterozygous sites divided by all sites.

626 # Treemix

627 Using the SAF files based on the data mapped to the goat reference genome, we called the
628 maximum-likelihood alternative allele count for each site found in every population, leading
629 to approximately 6.8 million input sites for TreeMix (Pickrell & Pritchard, 2012). For each
630 number of migrations between zero and five, we ran TreeMix 25 times with different seeds
631 and chose the best run for each possible migration. In all cases, we specified Etosha as the
632 root.

633 # EEMS

634 We used the bed2diffs function, as implemented in Estimating Effective Migration Surfaces
635 (EEMS) (Petkova et al., 2016), to generate an average genetic dissimilarity matrix based on
636 the imputed genotypes of dataset mapped to the goat reference genome, and for each
637 sample set as their coordinates the latitude and longitude of their sampling location
638 (Supplementary Table S1). Then we ran EEMS (Petkova et al., 2016) to assess the decay of
639 genetic similarity in a geospatial context, using 40 million iterations and a burnin of 20 million,
640 and 500 spatial demes spanning East and Southern Africa. We did three independent runs

641 and assessed convergence visually and with the Gelman–Rubin diagnostic in the R package
642 coda (Plummer et al., 2006).

643 D-statistics and f-branch

644 To understand deviations from treeness we summarised it as f-branch statistics using Dsuite (Malinsky
645 et al., 2018, 2021). This was done using the called genotypes from the medium-high depth individuals
646 based on data mapped to the goat reference genome. We use the tree inferred from TreeMix without
647 any migrations.
648
649 We also calculated D-statistics with the black-faced impala as P3 and the common impala as P1/P2
650 using the qpDstat from AdmixTools (Patterson et al., 2012) and summarised the results in a heatmap.
651 We used the called genotypes for the medium-high depth individuals, converted to binary PLINK
652 format using PLINK v1.9 (Chang et al., 2015).

653 Admixture graphs

654 To allow for a more detailed insight into the evolutionary history of impala, ADMIXTOOLS2 (Maier et
655 al., 2023) was used to estimate admixture graphs between populations. f2-scores were first extracted
656 from the PLINK file used to calculate D-statistics (see above) using the `extract_f2` function, using a
657 block length of 4 x 10^6 cM and retaining only SNPs with a maximal missingness of 0.9 within any given
658 population. The `find_graphs` function was run 500 times to heuristically search for likely graphs
659 describing admixture relationships between populations. In this process, a test score was calculated
660 by optimising a topology from a subset of the f-statistics and comparing all remaining graphs to this
661 topology. Significance testing was then performed using a jackknife approach, only retaining graphs
662 that were non-significantly worse than the graphs with the lowest test score (α=0.05). Graphs that
663 exhibited two admixture events from the same node or a ghost population were subsequently
664 discarded as these could not be distinguished from temporally implausible scenarios. Finally, the
665 number of allowed admixture events in the graph was incremented from zero to the highest number
666 of admixture events where the best test score was not significantly different from the best test score
667 of a less parameterised graph.

mtDNA analyses

mtDNA assembly

We mapped the samples to the impala mitochondrial genome (NCBI accession NC_020675, (Hassanin et al., 2012)), adding also a wildebeest sample (SRA BioSample ID SAMN39917065, (Liu et al., 2024)). All impala and wildebeest samples were mapped to the mitochondrial genome using the same pipeline as for the mapping to the whole genome with bwa mem. We then used ANGSD to produce a consensus fasta sequence for each of the samples, selecting the most common base in each position (-doFasta 2), and removing sites where more than 5 % of the reads supported a base different from the most common one.

mtDNA tree

The consensus sequences were aligned together with the wildebeest mitogenome sequence used as an outgroup. We used ModelTest-NG v0.1.7 (Darriba et al., 2020) to identify the best-fitting substitution scheme based on the Bayesian Information Criterion. HKY+I+G4 was selected, and with the gamma shape parameter of 0.8698 and the proportion of invariable sites of 0.7043 were used as priors in a phylogenetic tree reconstruction in BEAST v2.7 (Bouckaert et al., 2019). We used a strict molecular clock with a rate parameter of 3.95×10^{-8} (Colli et al., 2015) and the Yule tree model. The analysis was run with a MCMC chain length of 10,000,000 samples, logging samples every 1,000 steps and with a pre-burnin of 1,000. We used Tracer v1.7.1 (Rambaut et al., 2018) to verify that the MCMC trace showed convergence and all effective sample sizes had values above 200. TreeAnnotator was used to generate a maximum clade credibility tree based on the common ancestor heights, a 30% burn-in and a 0.7 posterior probability limit.

690 # Generation time

691 We used two different generation times, the first is the generation time given by the IUCN
692 (IUCN SSC Antelope Specialist Group, 2016) that was used in the Ruminant Genome Project
693 of (Chen et al., 2019), T = 5.7y. However, due to its substantial difference from another
694 published generation time for the impala of 7.3y (Pacifici et al., 2014), we decided to re-
695 estimate the generation time from demographic data. We used the survival rate from
696 (Spinage, 1972). The vector with the survival l_t until year t is

697 L_t = (1, 0.805, 0.753, 0.728, 0.650, 0.570, 0.372, 0.234, 0.130, 0.095, 0.061, 0.026).

698 The fecundity vector

699 M_t = (0, 0.25, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)

700 was based on the assumption that the sex ratio in the impala is 1 : 1 and on information
701 obtained from the AnAge Database of Animal Ageing and Longevity (Tacutu et al., 2018)
702 found at
703 https://genomics.senescence.info/species/entry.php?species=Aepyceros_melampus. The
704 generation time was estimated as
705

$$T = \sum_{t=\alpha}^{\omega} t l_t m_t / \left(\sum_{t=\alpha}^{\omega} l_t m_t \right) = 4.03,$$

706 where α is the age of the onset of reproduction and ω is the age of the last reproduction.

707 # Inference of long-term population size trajectories with
708 PSMC

709 We used the Pairwise Sequentially Markovian Coalescent (PSMC, (Li & Durbin, 2011)) to
710 estimate population size trajectories through time for the medium-high depth samples. To
711 have a very high depth sample, we used a subset of the short read data from the sample used

712 for the impala genome assembly ((Chen et al., 2019); see Sequencing in Methods section
713 above). We mapped and processed this sample with the sample pipeline as the rest.
714
715
716 Based on the data mapped to the impala genome, we called genotypes separately for each
717 sample with bcftools 1.10, using the consensus caller. We removed sites with a depth lower
718 than either ⅓ of the average sample sequencing depth, or lower than 6 sites if ⅓ of the
719 average was below 6. This minimum depth filter setting was based on an investigation of
720 different settings using subsampled versions of the high depth RGP sample (Supplementary
721 Figure S11B and S11C). We also removed sites with a depth higher than 2 times the average
722 sample depth, and heterozygous sites where there were less than 2 reads supporting either
723 of the two alleles. We then generated psmc input files and ran psmc with default settings. We
724 furthermore performed bootstrapping, by running psmc on bootstrapping mode without
725 splitting, since the scaffolds in the reference genome were already fragmented enough. We
726 scaled the output to years and diploid population sizes assuming a mutation rate of 1.41e-08
727 per site per generation (Chen et al., 2019) and two different generation times of 5.7 and 4.03
728 years per generation (see above Generation time section in Methods).

729 Inference of joint demographic history with fastsimcoal2

730 The demographic history of a representative subset of the impala populations (Etosha, Chobe,
731 Shangani and Masai Mara) was further investigated using a coalescent simulation- based
732 method implemented in Fastsimcoal2 v2.7.0.2 (Excoffier et al. 2013). To minimize potential
733 bias arising when determining ancestral allelic states, we used the folded 2dSFS, based on the
734 whole genome 2DSFS estimated with winsfs from the data mapped to the impala reference
735 genome, as previously described. Four plausible demographic models were tested
736 (Supplementary Figure. S12). As the null model, the Model 1 is a scenario without any
737 admixture/migration events and follows a population tree equivalent to the TreeMix tree
738 without any migration edges. The Model 2 scenario adds three admixture events with flexible
739 admixture proportions: one admixture event from Chobe to Etosha and two admixture events
740 happened at the same time before the split between Masai Mara and the ancestor of Chobe
741 and Shangani. The Model 3 models a scenario with one admixture event from Chobe to Etosha

742 and the asymmetric continuous migration between Shangani and Masai Mara. We also
743 considered a Model 4 similar to Model 3, but where asymmetric migration happened
744 continuously before the split between Masai Mara and the ancestor of Chobe and Shangani.
745
746 For each model we ran 100 independent Fastsimcoal runs to find the best-fitting parameters
747 yielding the highest likelihood, with 500,000 coalescent simulations per likelihood estimation
748 (-n500000), 100 conditional maximization algorithm cycles (-L100), and minimum 100
749 observed SFS entry count taken into account in likelihood computation (-C100). A mutation
750 rate of 1.41×10^{-8} per site per generation (Chen et al., 2019) and a generation time of 5.7 years
751 (IUCN SSC Antelope Specialist Group, 2016) and 4.03 years (see above section Generation
752 time in Methods) were used to convert model estimates from coalescence units to absolute
753 values (i.e. years).
754
755 For the two best models (Model 2 and Model 4), we estimated uncertainty using a non-
756 parametric jackknife approach. We used winsfs to estimate SFS in contiguous blocks of the
757 genome (Rasmussen et al., 2024), producing in total 50 blocks. We then produced 50 leave
758 one out SFS splits by summing the estimated SFS of all except one block each, and estimated
759 again the maximum likelihood parameters for each model, using the maximum likelihood
760 parameters inferred with the full data as initial guesses. For each split we fitted 5 independent
761 runs, and selected the maximum likelihood parameters for each case. Using the 50 maximum
762 likelihood parameters from each split, we estimated standard errors for each parameter using
763 the block jackknife estimator (Busing et al., 1999).

764 Runs of homozygosity

765 We estimated Runs of Homozygosity (ROH) for all samples using the imputed dataset mapped
766 to the goat genome, since the impala reference genome was too fragmented to identify ROHs.
767 We used PLINK 1.9 to call regions in ROH, and after testing different settings used the
768 following: minimum length to call a ROH of 500 kbp (--homozyg-kb 500), minimum SNP
769 density within ROH of 1 SNP per 100 kbp (--homozyg-density 100), used scanning windows of
770 50 SNPs (--homozyg-window-snp 50), allowing at most 5 heterozygous call within a window
771 in ROH (--homozyg-kb 500), allowing 20 missing SNPs per window in ROH (--homozyg-

window-missing 20), and using otherwise default settings. Finally, we manually merged ROHs within 500 kb of each other, after observing a high proportion of ROH within a very short distance most likely due to fragmentation of ROH due to excess missigness or genotyping errors. These settings are quite relaxed, but this was necessary due to a high proportion of spurious heterozygous calls that are unavoidable when mapping to a distant reference genome, which has been shown to often prevent the identification of ROHs (Prasad et al., 2022). Moreover, most of the samples are low depth and imputed without a reference panel, introducing further uncertainty in the genotype calls. However, these limitations affect all the populations equally, and thus, while we advise caution when interpreting the absolute estimated proportion of the genome in ROH for each category, the relative amount of ROHs identified per population can be considered reliable .

LD-decay

Pairwise LD between sites within populations was estimated using the R package https://github.com/aalbrechtsen/relate (Albrechtsen et al., 2009) and summaries into curves in R as described in https://github.com/aalbrechtsen/LDdecay. We used imputed genotypes mapped to the goat reference genome as input used after removing sites with a minor allele frequency below 5%.

Efficiency of selection

We used the amount of derived alleles in positions with putatively different functional effects as a proxy for the efficiency of selection. Using the publicly available goat genome annotation, we extracted positions from coding sequences where mutations would change the amino acid sequence of the encoded protein (0-fold degenerate sites), and from those where nucleotide mutations would not alter the amino-acid sequence (4-fold degenerate sites). We assume that mutations in 0-fold degenerate sites would predominantly be neutral or deleterious, while mutations at 4-fold degenerate sites would always be neutral (or would at least be neutral much more often than at 0-fold sites), which is a common assumption made in studies of genetic load and efficiency of selection (Bertorelle et al., 2022). We then estimated population level SAF files from the data mapped to the goat reference genome, separately for each of these two positions using ANGSD with the same settings as previously described,

using the allele in the goat genome to assign ancestral state. From the SAF files we estimated the SFS for each population and position type with realSFS. Finally, we used as a proxy of the efficiency of selection within each population, the ratio of the total proportion of derived alleles found within the population at 0-fold degenerate sites to the total proportion of derived alleles found within the population at 4-fold degenerate sites.

References

Albrechtsen, A., Sand Korneliussen, T., Moltke, I., van Overseem Hansen, T., Nielsen, F. C., & Nielsen, R. (2009). Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genetic Epidemiology*, *33*(3), 266–274.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arctander, P., Johansen, C., & Coutellec-Vreto, M. A. (1999). Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution*, *16*(12), 1724–1739.

Bastos-Silveira, C., & Lister, A. M. (2007). A morphometric assessment of geographical variation and subspecies in impala. *Journal of Zoology*, *271*(3), 288–301.

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews. Genetics*, *23*(8), 492–503.

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Research*, *23*(9), 1514–1521.

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T., Burton, J. N., Huson, H. J., Nystrom, J. C., Kelley, C. M., Hutchison, J. L., Zhou, Y., Sun, J., Crisà, A., Ponce de León, F. A., … Smith, T. P. L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, *49*(4), 643–650.

826 Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled,
827 J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H.
828 A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J.
829 (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS*
830 *Computational Biology*, *15*(4), e1006650.
831 Browning, B. L., & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-
832 phase inference for large data sets of trios and unrelated individuals. *American Journal of*
833 *Human Genetics*, *84*(2), 210–223.
834 Busing, F. M. T. A., Meijer, E., & Leeden, R. V. D. (1999). Delete-m Jackknife for Unequal m. *Statistics*
835 *and Computing*, *9*(1), 3–8.
836 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-
837 generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.
838 Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., Su, W., Liu, G.,
839 Li, Q., Fu, W., Pan, X., Liu, C., Yang, J., Zhang, C., Yin, Y., … Wang, W. (2019). Large-scale
840 ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*,
841 *364*(6446). https://doi.org/10.1126/science.aav6202
842 Clemente, F., & Vogl, C. (2012). Evidence for complex selection on four-fold degenerate sites in
843 Drosophila melanogaster. *Journal of Evolutionary Biology*, *25*(12), 2582–2595.
844 Colli, L., Lancioni, H., Cardinali, I., Olivieri, A., Capodiferro, M. R., Pellecchia, M., Rzepus, M., Zamani,
845 W., Naderi, S., Gandini, F., Vahidi, S. M. F., Agha, S., Randi, E., Battaglia, V., Sardina, M. T.,
846 Portolano, B., Rezaei, H. R., Lymberakis, P., Boyer, F., … Achilli, A. (2015). Whole mitochondrial
847 genomes unveil the impact of domestication on goat matrilineal variability. *BMC Genomics*, *16*,
848 1115.
849 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T.,
850 McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools.
851 *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, *37*(1), 291–294.

East, R. (1998). *African Antelope Database*. IUCN/SSC Antelope Specialist Group.

Eldredge, N., & Gould, S. J. (1972). Punctuated equilibria: An alternative to phyleytic gradualism. In T. J. M. Schopf (Ed.), *Models in Paleobiology* (pp. 82–115). Freeman, Cooper and Company.

Eldredge, N., Thompson, J. N., Brakefield, P. M., Gavrilets, S., Jablonski, D., Jackson, J. B. C., Lenski, R. E., Lieberman, B. S., McPeek, M. A., & Miller, W. (2005). The dynamics of evolutionary stasis. In *Paleobiology* (Vol. 31, Issue sp5, pp. 133–145). https://doi.org/10.1666/0094-8373(2005)031[0133:tdoes]2.0.co;2

Estes, R. (1991). *The Behavior Guide to African Mammals*. University of California Press.

Estes, R. (1993). *The Safari Companion: A Guide to Watching African Mammals : Including Hoofed Mammals, Carnivores, and Primates*. Chelsea Green Publishing Company.

Estes, S., & Arnold, S. J. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, *169*(2), 227–244.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* , *32*(19), 3047–3048.

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* , *37*(24), 4882–4885.

Faith, J. T., Tyler Faith, J., Tryon, C. A., Peppe, D. J., Beverly, E. J., & Blegen, N. (2014). Biogeographic and Evolutionary Implications of an Extinct Late Pleistocene Impala from the Lake Victoria Basin, Kenya. In *Journal of Mammalian Evolution* (Vol. 21, Issue 2, pp. 213–222). https://doi.org/10.1007/s10914-013-9238-1

Fritz, H., & Bourgarel, M. (2013). *Aepyceros melampus* Impala. In J. Kingdon & M. Hoffman (Eds.), *Mammals of Africa* (Vol. 6, pp. 480–488). Bloomsbury Publishing.

Garcia-Erill, G., & Albrechtsen, A. (2020). Evaluation of model fit of inferred admixture proportions. *Molecular Ecology Resources*, *20*(4), 936–949.

Garcia-Erill, G., Jørgensen, C. H. F., Muwanika, V. B., Wang, X., Rasmussen, M. S., de Jong, Y. A., Gaubert, P., Olayemi, A., Salmona, J., Butynski, T. M., Bertola, L. D., Siegismund, H. R., Albrechtsen, A., & Heller, R. (2022). Warthog Genomes Resolve an Evolutionary Conundrum and Reveal Introgression of Disease Resistance Genes. *Molecular Biology and Evolution*, *39*(7). https://doi.org/10.1093/molbev/msac134

Gaspar, J. M. (2018). NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*, *19*(1), 536.

Gentry, A. W. (1985). The Bovidae of the Omo deposits, Ethiopia. In Coppens Y And Howell F (Ed.), *Les faunes Plio-Pléistocènes de la basse Vallée de l'Omo (Ethiopie); I: Perissodactyles-Artio-dactyles (Bovidae)* (pp. 119–191). Cahiers de Paleontologie.

Geraads, D. (2019). A reassessment of the Bovidae (Mammalia) from the Nawata Formation of Lothagam, Kenya, and the late Miocene diversification of the family in Africa. *Journal of Systematic Palaeontology*, *17*(2), 169–182.

Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, *3*(2), 115–151.

Green, W. C. H., & Rothstein, A. (1998). Translocation, Hybridization, and the Endangered Black-Faced Impala. *Conservation Biology: The Journal of the Society for Conservation Biology*, *12*(2), 475–480.

Groves, C., & Grubb, P. (2011). *Ungulate Taxonomy*. JHU Press.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T. T., & Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies*, *335*(1), 32–50.

Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2019). The great human expansion. In

Resonance (Vol. 24, Issue 6, pp. 711–718). https://doi.org/10.1007/s12045-019-0830-4

Hunt, G., & Rabosky, D. L. (2014). Phenotypic Evolution in Fossil Species: Pattern and Process. *Annual Review of Earth and Planetary Sciences*, *42*(1), 421–441.

IUCN SSC Antelope Specialist Group. (2016). *Aepyceros melampus*. *The IUCN Red List of Threatened Species*, *e. T550A50180828*.

Joubert, E. (1971). Observations on the habitat preferences and population dynamics of the black-faced Impala *Aepyceros petersi* Bocage, 1875, in South West Africa. *Madoqua*, *I*(3), 55–65.

Kingdon, J. (2013). GENUS Aepyceros. In J. Kingdon & M. Hoffman (Eds.), *Mammals of Africa* (Vol. 6, p. 479). Bloomsbury Publishing.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356.

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. https://doi.org/10.48550/ARXIV.1303.3997

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

Liu, X., Lin, L., Sinding, M.-H. S., Bertola, L. D., Hanghøj, K., Quinn, L., Garcia-Erill, G., Rasmussen, M. S., Schubert, M., Pečnerová, P., Balboa, R. F., Li, Z., Heaton, M. P., Smith, T. P. L., Pinto, R. R., Wang, X., Kuja, J., Brüniche-Olsen, A., Meisner, J., … Heller, R. (2024). Introgression and disruption of migration routes have shaped the genetic integrity of wildebeest populations. *Nature Communications*, *15*(1), 2921.

Lorenzen, E. D., Arctander, P., & Siegismund, H. R. (2006). Regional Genetic Structuring and Evolutionary History of the Impala *Aepyceros melampus*. In *Journal of Heredity* (Vol. 97, Issue 2, pp. 119–132). https://doi.org/10.1093/jhered/esj012

Lorenzen, E. D., Heller, R., & Siegismund, H. R. (2012). Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, *21*(15), 3656–3670.

Lorenzen, E. D., Masembe, C., Arctander, P., & Siegismund, H. R. (2010). A long-standing Pleistocene

refugium in southern Africa and a mosaic of refugia in East Africa: insights from mtDNA and the common eland antelope. *Journal of Biogeography*, *37*(3), 571–581.

Lorenzen, E. D., & Siegismund, H. R. (2004). No suggestion of hybridization between the vulnerable black-faced impala (Aepyceros melampus petersi) and the common impala (A. m. melampus) in Etosha National Park, Namibia. In *Molecular Ecology* (Vol. 13, Issue 10, pp. 3007–3019). https://doi.org/10.1111/j.1365-294x.2004.02308.x

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, *22*(5), 1678–1692.

Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., & Reich, D. (2023). On the limits of fitting complex models of population history to -statistics. *eLife*, *12*. https://doi.org/10.7554/eLife.85492

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595.

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955.

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

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731.

Meisner, J., & Albrechtsen, A. (2019). Testing for Hardy–Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. In *Molecular Ecology Resources* (Vol. 19, Issue 5, pp. 1144–1152). https://doi.org/10.1111/1755-0998.13019

Miller, S. M., Moeller, C.-H., Harper, C. K., & Bloomer, P. (2020). Anthropogenic movement results in hybridisation in impala in southern Africa. In *Conservation Genetics* (Vol. 21, Issue 4, pp. 653–663). https://doi.org/10.1007/s10592-020-01276-4

Nersting, L. G., & Arctander, P. (2001). Phylogeography and conservation of impala and greater kudu. *Molecular Ecology*, *10*(3), 711–719.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PloS One*, *7*(7), e37558.

Nursyifa, C., Brüniche-Olsen, A., Garcia-Erill, G., Heller, R., & Albrechtsen, A. (2022). Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Molecular Ecology Resources*, *22*(2), 458–467.

Olson, M. E. (2024). Is Population Genetics Really Relevant to Evolutionary Biology? *Evolutionary Biology*, *51*(2), 235–243.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Grottolo Marasini, G., Visconti, P., & Rondinini, C. (2014). *Data from: Generation length for mammals* [Dataset]. Dryad. https://doi.org/10.5061/DRYAD.GD0M3

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093.

Pečnerová, P., Garcia-Erill, G., Liu, X., Nursyifa, C., Waples, R. K., Santander, C. G., Quinn, L., Frandsen, P., Meisner, J., Stæger, F. F., Rasmussen, M. S., Brüniche-Olsen, A., Hviid Friis Jørgensen, C., da Fonseca, R. R., Siegismund, H. R., Albrechtsen, A., Heller, R., Moltke, I., &

Hanghøj, K. (2021). High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Current Biology: CB*, *31*(9), 1862–1871.e5.

Pedersen, C.-E. T., Albrechtsen, A., Etter, P. D., Johnson, E. A., Orlando, L., Chikhi, L., Siegismund, H. R., & Heller, R. (2018). A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology & Evolution*, *2*(3), 491–498.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

Pockrandt, C., Alzamel, M., Iliopoulos, C. S., & Reinert, K. (2020). GenMap: ultra-fast computation of genome mappability. *Bioinformatics* , *36*(12), 3687–3692.

Prasad, A., Lorenzen, E. D., & Westbury, M. V. (2022). Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. *Molecular Ecology Resources*, *22*(1), 45–55.

Quinn, L., Garcia-Erill, G., Santander, C., Brüniche-Olsen, A., Liu, X., Sinding, M.-H. S., Heaton, M. P., Smith, T. P. L., Pečnerová, P., Bertola, L. D., Hanghøj, K., Rasmussen, M. S., de Jager, D., Siegismund, H. R., Albrechtsen, A., Heller, R., & Moltke, I. (2023). Colonialism in South Africa leaves a lasting legacy of reduced genetic diversity in Cape buffalo. *Molecular Ecology*. https://doi.org/10.1111/mec.16851

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Rasmussen, M. S., C., Wiuf, & Albrechtsen, A. (2024). Inferring drift, genetic differentiation, and admixture graphs from low-depth sequencing data. *bioRxiv*. https://doi.org/10.1101/2024.01.29.577762

Rasmussen, M. S., Garcia-Erill, G., Korneliussen, T. S., Wiuf, C., & Albrechtsen, A. (2022). Estimation

of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage. *Genetics*, *222*(4). https://doi.org/10.1093/genetics/iyac148

Reynolds, S. C. (2007). Mammalian body size changes and Plio-Pleistocene environmental shifts: implications for understanding hominin evolution in eastern and southern Africa. *Journal of Human Evolution*, *53*(5), 528–548.

Reynolds, S. C. (2010). Morphological evaluation of genetic evidence for a Pleistocene extirpation of eastern African impala. *South African Journal of Science*, *106*(11), 1–7.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.

Spinage, C. A. (1972). African ungulate life tables. *Ecology*, *53*(4), 645–652.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., & de Magalhães, J. P. (2018). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, *46*(D1), D1083–D1090.

Voje, K. L. (2018). Assessing adequacy of models of phyletic evolution in the fossil record. *Methods in Ecology and Evolution / British Ecological Society*, *9*(12), 2402–2413.

Vrba, E. S. (1980). Evolution, species and fossils: how does life evolve? *South African Journal of Science*, *76*(2), 61–84.

Vrba, E. S. (1984). Evolutionary Pattern and Process in the Sister-Group Alcelaphini-Aepycerotini (Mammalia: Bovidae). In N. Eldredge & S. M. Stanley (Eds.), *Living Fossils* (pp. 62–79). Springer.

Wang, X., Pedersen, C.-E. T., Athanasiadis, G., Garcia-Erill, G., Hanghøj, K., Bertola, L. D., Rasmussen, M. S., Schubert, M., Liu, X., Li, Z., Lin, L., Jørsboe, E., Nursyifa, C., Liu, S., Muwanika, V., Masembe, C., Chen, L., Wang, W., Moltke, I., … Heller, R. (2022). Persistent gene flow suggests an absence of reproductive isolation in an African antelope speciation model. In *bioRxiv* (p. 2022.12.08.519574). https://doi.org/10.1101/2022.12.08.519574

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial

relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*(1), 35–48.

Data availability

All raw sequencing data generated for this study is available in SRA under BioProject accession PRJNA862915. Code used for analyses and visualisations is available at https://github.com/GenisGE/impalascripts.

Acknowledgements

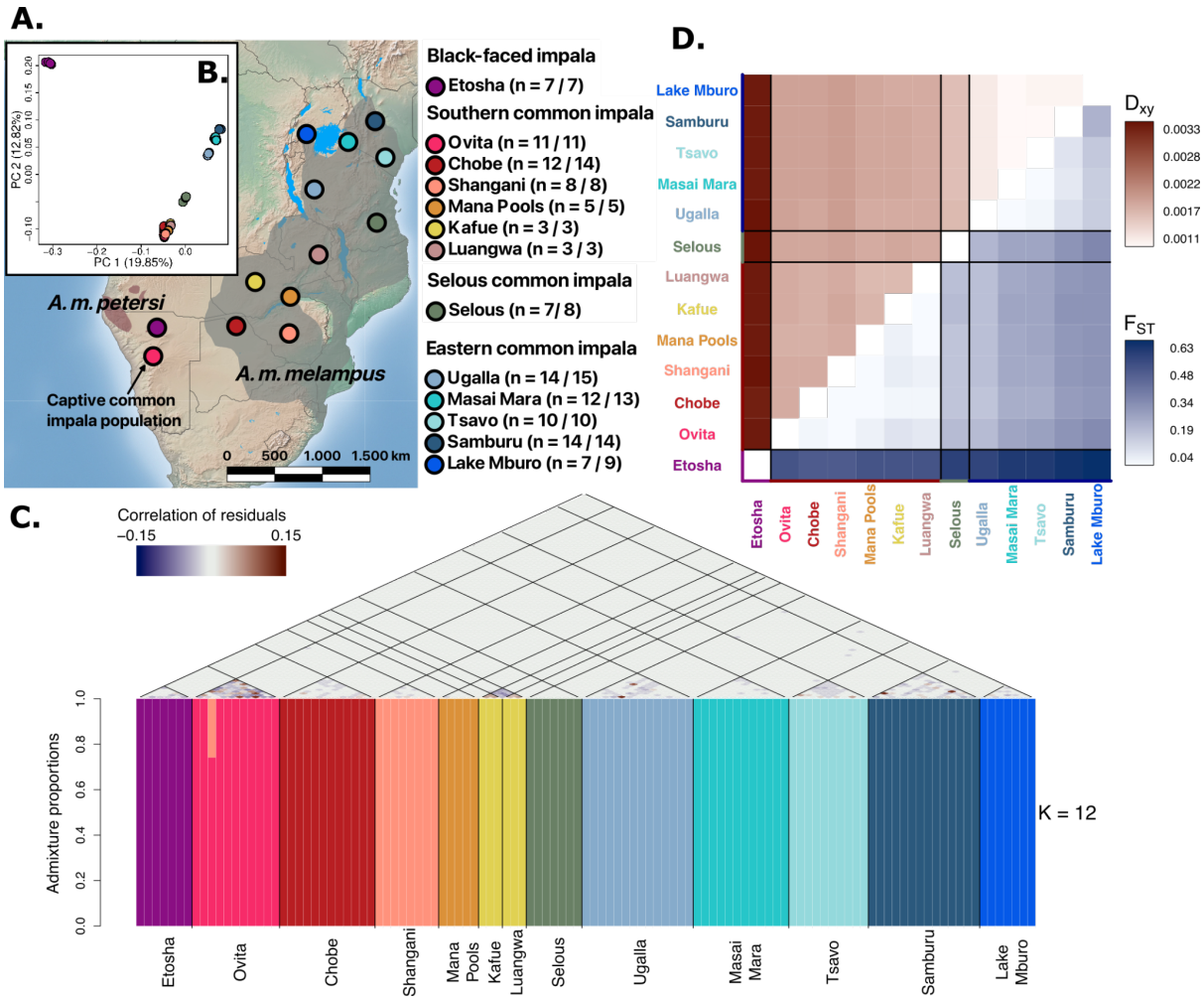
We thank Amal Al-Chaer for her work on extracting DNA for this study, Peter Arctander for organising the collection of samples, David Moyer for contributing samples to the collection and the African wildlife management authorities for giving permissions and collaborating with Peter Arctander during sample collection. GGE and RH were supported by a Danmarks Frie Forskningsfond Sapere Aude research grant (DFF8049-00098B), RH was further supported by a Carlsberg Young Researcher grant (CF21-0497) and AA, GGE, MSR and KH were supported by Danmarks Frie Forskningsfond (DFF-0135-00211B). We thank the journal editor and two anonymous reviewers for their suggestions that helped improve this manuscript.

Author contributions

GGE: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing-original draft, Writing-review & editing; XW: Formal Analysis, Visualization, Writing-review & editing; MSR: Formal Analysis, Methodology, Software, Writing-review & editing; LQ: Conceptualization, Formal Analysis, Writing-review & editing; AK: Conceptualization, Writing-original draft, Writing-review & editing; LDB: Formal Analysis, Visualization, Writing-review & editing; CGG: Formal Analysis, Writing-review & editing; RFB: Formal Analysis, Writing-review & editing; JOO: Conceptualization, Writing-review & editing; PP: Formal Analysis, Writing-review & editing; KH: Methodology, Software, Writing-review & editing; JK: Formal Analysis,

Visualization, Writing-review & editing; CN: Formal Analysis, Writing-review & editing; CM: Data curation, Resources, Writing-review & editing; VM: Data curation, Resources, Writing-review & editing; FB: Conceptualization, Writing-review & editing; IM: Conceptualization, Writing-review & editing; HRS: Conceptualization, Data Curation, Resources, Writing-review & editing; AA: Conceptualization, Methodology, Funding acquisition, Project administration, Supervision, Writing-review & editing; RH: Conceptualization, Data Curation, Funding acquisition, Methodology, Project administration, Supervision, Writing-original draft, Writing-review & editing

Figures



1071 **Figure 1: Distribution and population structure of impalas**. A. Distribution map and sample
1072 localities of impala. Shaded regions indicate the current distribution range of each impala
1073 subspecies (IUCN SSC Antelope Specialist Group, 2016). The sample sizes indicated in the
1074 legend correspond to the 113 samples kept in the filtered dataset (before the slash) and the
1075 initial set of 119 samples sequenced (after the slash), with each locality having at least 1
1076 medium-high depth sample. B. Principal component analysis performed by applying PCAngsd
1077 to genotype likelihoods from common SNPs across all impala populations. C. Admixture
1078 proportions estimated with NGSadmix, assuming 12 ancestral populations (K=12, lower
1079 panel), together with an evaluation of model fit through the correlation of residuals for each
1080 pair of individuals estimated with evalAdmix (upper panel). D. Genetic divergence and
1081 population differentiation between each population pair, measured respectively with Dxy and
1082 with F_{ST} estimated using Hudson's estimator. Both quantities were estimated from the
1083 inferred 2-dimensional site frequency spectrum (2DSFS).
1084
1085

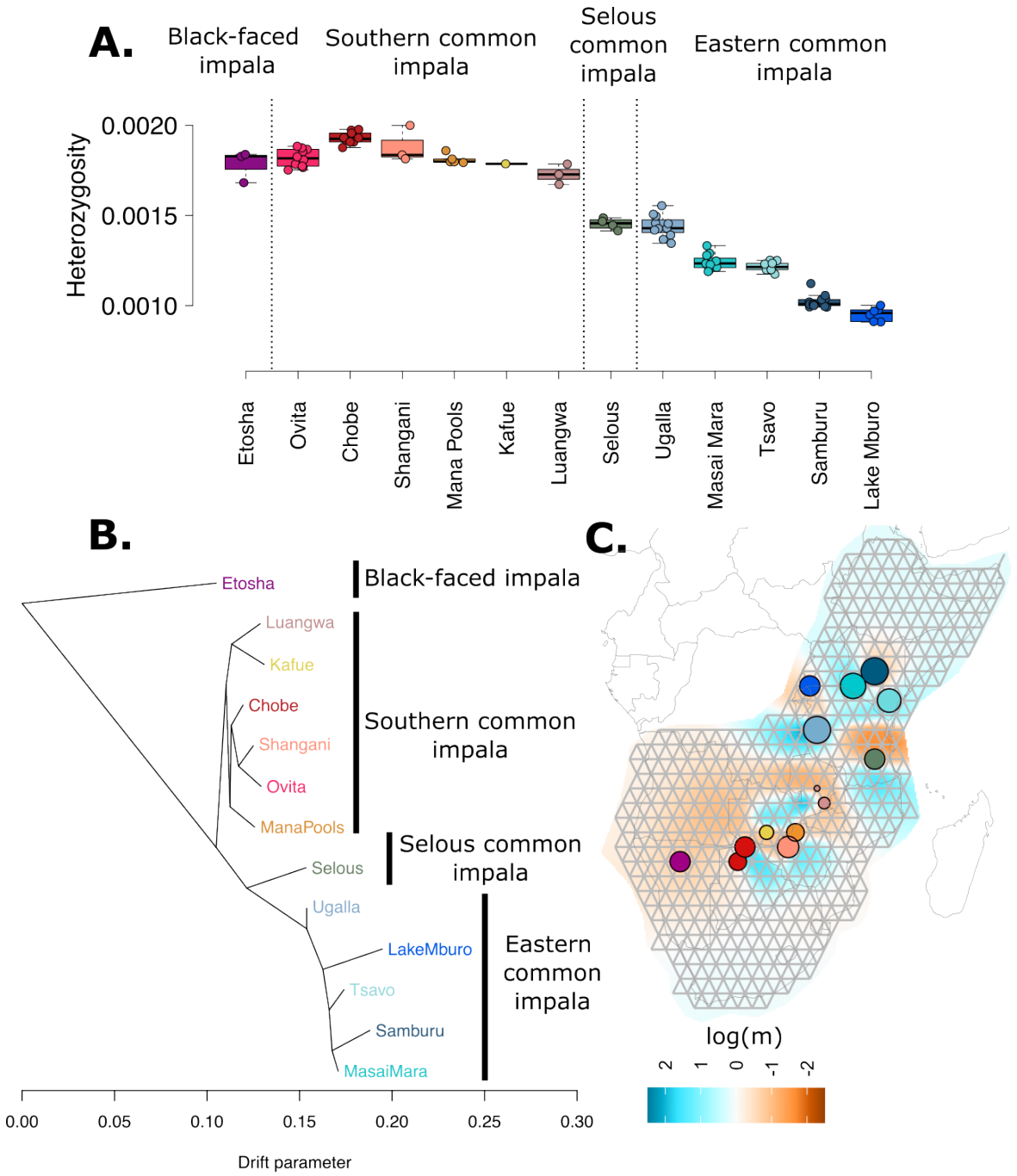


Figure 2: Genetic diversity and phylogeography of impala: A. Per-individual heterozygosity stratified by sampling location for all the impala samples. To minimise bias caused by sequencing errors, we excluded low-depth samples with a sequencing error rate above 0.02% (Supplementary Figure S6). B. Population tree estimated with TreeMix without allowing for any migration events. Labels on the right side show the three main geographical groupings arising from the data. C. Effective migration surface across the species range estimated with EEMS, using data from all the sampled impala populations except the captive common impala population from Ovita Farm. The log(m) values indicate relative migration rates, with lower

values (red) corresponding to reduced migration rates and higher values (blue) corresponding to elevated migration rates with respect to the average.

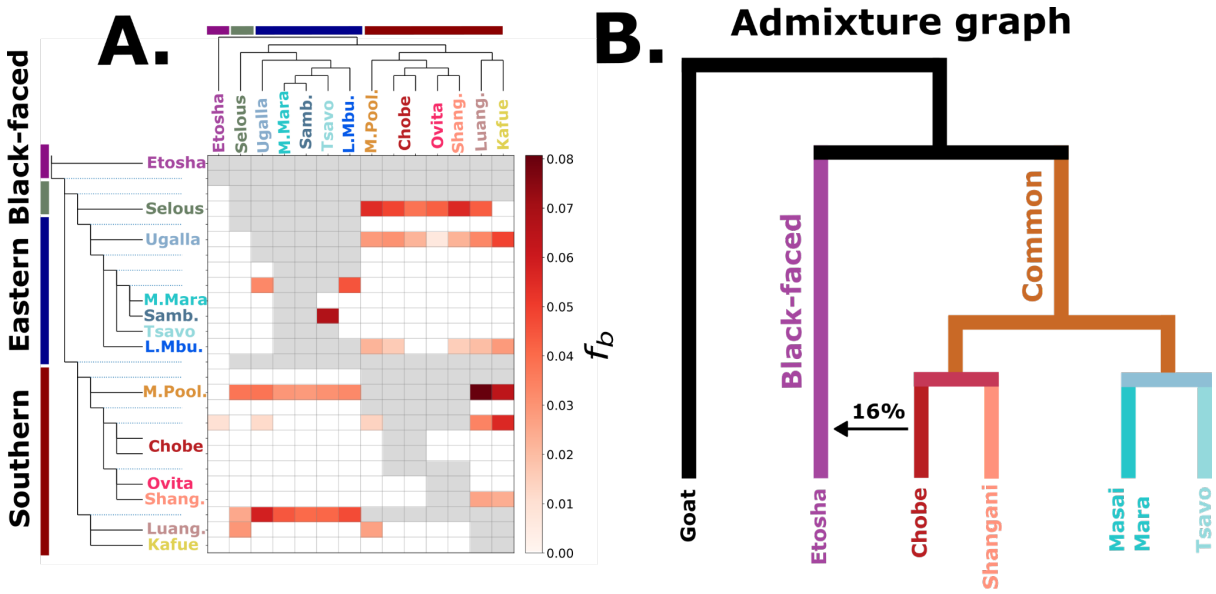


Figure 3: Gene flow between impala populations. A. Summary of all f4 statistics reflecting a significant violation of the population tree without migration (Figure 3B) obtained using the f-branch statistic. Each population is represented by their medium-high depth sample (two for Chobe). B. Admixture graph obtained by applying qpGraph to medium-high depth samples for a subset of the populations. Branches are not drawn to scale. The graph shows a possible scenario of introgression into the black-faced impala from Chobe, its neighbouring common impala population (see Supplementary Figure S10 for other admixture graphs that explain the data equally well).

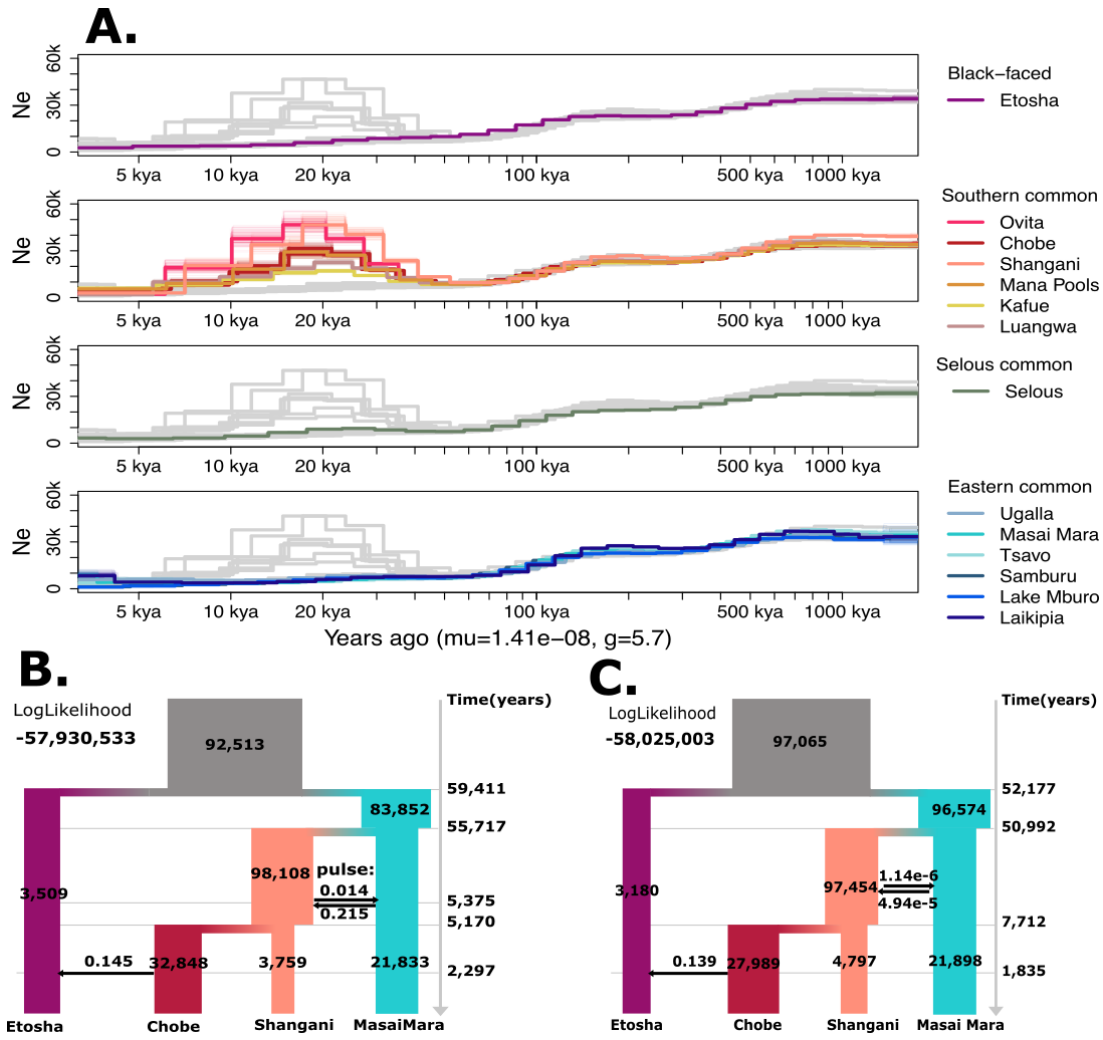


Figure 4: Demographic history of impala: A. Long-term effective population sizes for all impala populations, estimated using the medium-high depth samples and the Pairwise Sequentially Markovian Coalescent (PSMC) algorithm. Each panel highlights coloured samples from each of the three main groups, with bootstrap estimates shown in thinner lines, while the point estimates from non-highlighted samples in each corresponding panel are shown in grey. The high depth sample from the Ruminant Genome Project, collected in Laikipia, Kenya, is included for comparison. B. Joint demographic history for a subset of populations estimated from the 2DSFS with fastsimcoal2, including population split times and admixture events. Time and population sizes are not to scale and the log-likelihood is shown. C. Same model but assuming continuous gene flow between the population ancestral to Shangani and Chobe with Masai Mara.

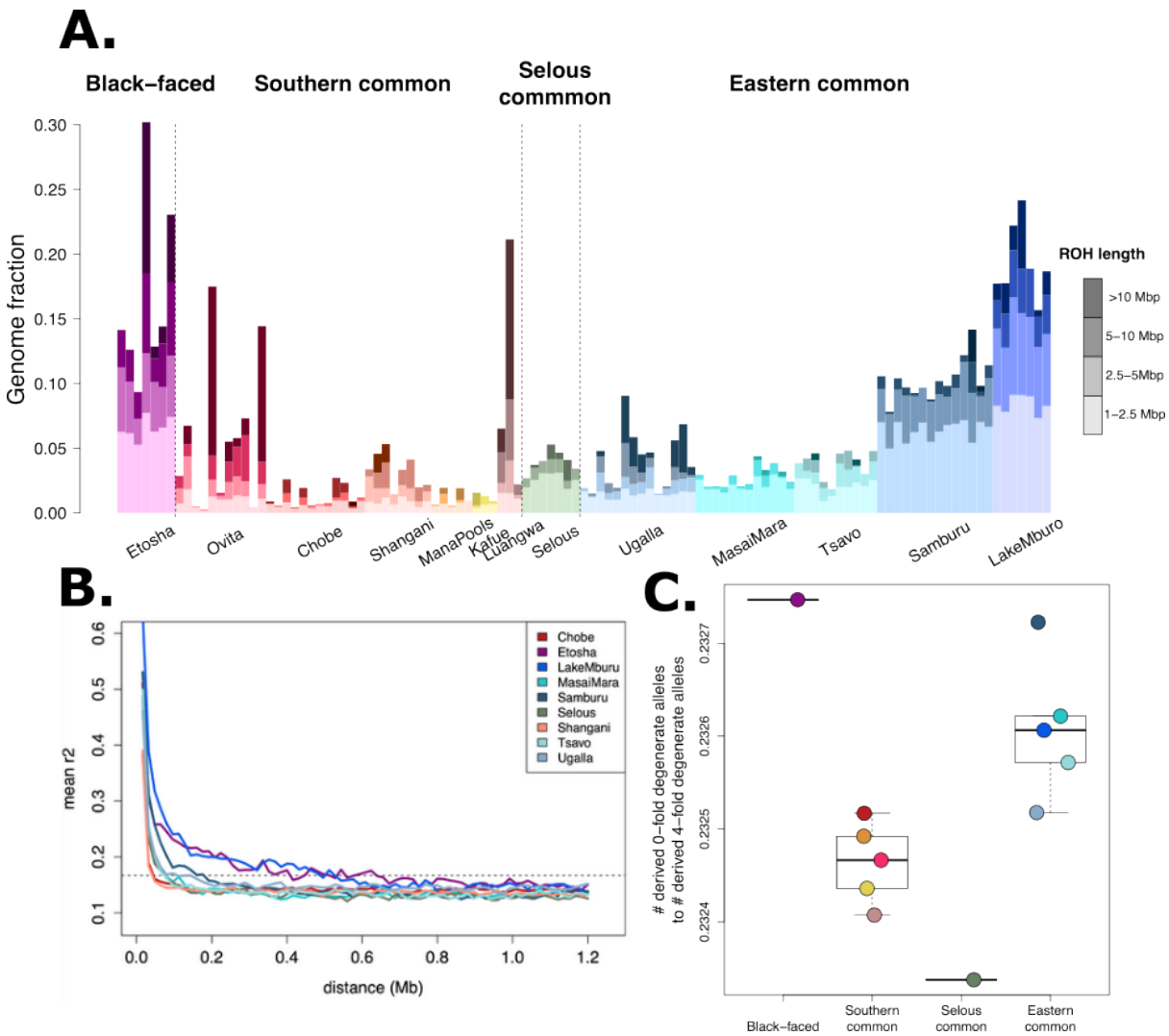


Figure 5: Impact of recent demography in impala genomes. A. Proportion of the genome for each sample covered by runs of homozygosity (ROH). The colour hue distinguishes populations and the colour brightness differentiates the ROH length bins. B. Linkage-disequilibrium decay curves for each population based on N=7 randomly selected individuals from each population. A horizontal dotted guide line is shown at 1/(N-1). C. The ratio of the total number of derived alleles at 0-fold degenerate sites to those at 4-fold degenerate sites within each impala population. The population from Shangani is omitted from this plot due to a likely excess of derived alleles caused by excessive error rates, which could impact the ratio (Supplementary Figure S16) .