Genomic analysis reveals limited hybridization among three giraffe species in Kenya

Raphael T. F. Coimbra[1,2,*], Sven Winter[1], Arthur Muneza[3], Stephanie Fennessy[3], Moses Otiende[4], Domnic Mijele[4], Symon Masiaine[5], Jenna Stacy-Dawes[5], Julian Fennessy[3,6], Axel Janke[1,2,7,*]

[1]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

[2]Institute for Ecology, Evolution and Diversity, Goethe University, Frankfurt am Main, Germany

[3]Giraffe Conservation Foundation, Windhoek, Namibia

[4]Kenya Wildlife Service, Nairobi, Kenya

[5]San Diego Zoo Wildlife Alliance, San Diego, USA

[6]School of Biology and Environmental Science, University College Dublin, Ireland

[7]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt am Main, Germany

*Correspondence: raphael.t.f.coimbra@gmail.com; axel.janke@senckenberg.de

Abstract

Background: In the speciation continuum the strength of reproductive isolation varies, and species boundaries are blurred by gene flow. Interbreeding among giraffe (*Giraffa* spp.) in captivity is known and anecdotal reports of natural hybrids exist. In Kenya, Nubian (*G. camelopardalis camelopardalis*), reticulated (*G. reticulata*), and Masai giraffe sensu stricto (*G. tippelskirchi tippelskirchi*) are parapatric, and thus the country might be a melting pot for these taxa. We analyzed 128 genomes of wild giraffe, 113 newly sequenced, representing these three taxa.

Results: We found varying levels of Nubian ancestry in 13 reticulated giraffe sampled across the Laikipia Plateau most likely reflecting historical gene flow between these two lineages. Although comparatively weaker signs of ancestral gene flow and potential mitochondrial introgression from reticulated into Masai giraffe were also detected, estimated admixture levels between these two lineages are minimal. Importantly, contemporary gene flow between East African giraffe lineages was not statistically significant. Effective population sizes have declined since the Late Pleistocene, more severely for Nubian and reticulated giraffe.

Conclusions: Despite historically hybridizing, these three giraffe lineages have maintained their overall genomic integrity suggesting effective reproductive isolation, consistent with the previous classification of giraffe into four species.

Keywords: East Africa, gene flow, *Giraffa*, hybridization, introgression, population genomics, speciation, whole-genome sequencing

Background

Speciation is a continuous process that can be thought of as a spectrum of reproductive isolation [1]. Depending on the strength of the reproductive barriers, hybridization may lead to introgression and gene flow in areas of range overlap [2]. Introgressive hybridization can homogenize the genomic landscape of incipient species until they break down into hybrid swarms [3]. Alternatively, it may also enhance evolutionary potential by increasing the frequency of favorable genetic variants or introducing novel allele combinations [4]. These processes can create phylogenetic incongruence across the genome resulting in reticulate evolution and blurring species boundaries [5, 6]. Mounting evidence demonstrates that natural hybridization and gene flow between related species are common [7], as has been observed, for instance, in *Heliconius* butterflies [8], Darwin's finches [9], and Grant's gazelles [10].

Speciation and the number of species in giraffe (*Giraffa* spp.) has gathered considerable interest in recent years [11–17]. Giraffe have a wide and fragmented distribution throughout sub-Saharan Africa [18]. They are capable of long-distance movements up to 300 km [19] and can have home ranges as large as 1,950 km^2 [20]. When housed together in captivity, some taxa can readily interbreed [21–23]. Yet, current taxonomic assessments based on nuclear and mitochondrial genetic data support either three [16] or four [17] highly divergent lineages with sub-structuring. Herein, we adopt the nomenclature used in Coimbra et al. [17], which includes four species and seven subspecies – the northern giraffe (*G. camelopardalis*), including West African (*G. c. peralta*), Kordofan (*G. c. antiquorum*), and Nubian giraffe (*G. c. camelopardalis* senior synonym of *G. c. rothschildi*); the reticulated giraffe (*G. reticulata*); the Masai giraffe sensu lato (*G. tippelskirchi*), including Luangwa (*G. t. thornicrofti*) and Masai giraffe sensu stricto (*G. t. tippelskirchi*); and, the southern giraffe (*G. giraffa*), including South African (*G. g. giraffa*) and Angolan giraffe (*G. g. angolensis*).

In East Africa, Nubian, reticulated, and Masai giraffe s. str. are largely parapatric [18] (Fig. 1a). In Kenya, their ranges adjoin in the seeming absence of natural barriers to dispersal in recent times [11, 24]. There have been anecdotal reports of individuals exhibiting intermediate phenotypes in the region [25–28], however, genetic admixture between giraffe species in the wild seems to be limited, implying that natural hybridization is rare [11, 14]. Reproductive asynchrony and seasonal variation in habitat use, both possibly related to regional differences in seasonality of rainfall and associated emergence and availability of browse, or pelage-based assortative mating may non-exclusively contribute to maintain genetic and phenotypic divergence (i.e., genetic structure in nuclear and mitochondrial DNA, and differences in pelage pattern) among these taxa [11, 24]. Nevertheless, hybridization between giraffe species in the wild has not been studied at a genomic scale.

Contact zones provide unique opportunities to understand the nature of species boundaries and the processes involved in the onset and maintenance of speciation [29]. Moreover, as modern genomics enhances our ability to uncover species divergence in the presence of gene flow [30], our perception of hybridization and its consequences for the conservation of biodiversity deepens [31]. Here, we investigate the extent of hybridization and genetic admixture among the three giraffe taxa occurring in East Africa, focusing specifically on Kenya, and reconstruct changes in their population size in the recent past. We analyzed the complete nuclear and mitochondrial genomes of 128 wild giraffe sampled mostly across Kenya, including from suspected contact zones of Nubian and reticulated giraffe (i.e., the Laikipia Plateau; Fig. 1a, locations 8–13), and of reticulated and Masai giraffe s. str. (i.e., south of Garissa towards the Tsavo Region; Fig. 1a, locations 17, 30, and 31) [26]. This first genome-scale assessment of hybridization among East Africa's giraffe lineages will aid to redefine their taxonomic status on the International Union for Conservation of Nature (IUCN) Red List and plan targeted conservation interventions for these threatened taxa [18].

Results

Genome resequencing. We analyzed genomes from 113 newly sequenced wild giraffe from across Kenya and 15 publicly available giraffe genomes from Ethiopia, Kenya, Tanzania, and Uganda (Fig. 1a and Additional file 1: Table S1), representing three separately evolving lineages: Nubian, reticulated, and Masai giraffe s. str. Read mapping against a chromosome-level Masai giraffe s. str. genome assembly [32] (GenBank: GCA_013496395) resulted in a mean mapping rate of 98.6% and a median filtered depth of 9× (1–26×) (Additional file 1: Table S1).

Population structure and admixture. After filtering our dataset against relatedness (Additional file 2: Fig. S1), a principal component analysis (PCA) and an admixture analysis

assuming three ancestry components (*K*) based on 484,876 unlinked single nucleotide polymorphisms (SNPs) correctly assigned all giraffe individuals to their respective species (Fig. 1). In the PCA (Fig. 1b), the first two principal components (PCs) explain most of the variance in the dataset, separating Nubian and reticulated from Masai giraffe s. str. (PC1: 72.46%), and Nubian from reticulated giraffe (PC2: 17.78%). On PC2, reticulated giraffe individuals from the Laikipia Plateau in Kenya (Fig. 1a, locations 8–13) are spread between Nubian and the remaining reticulated giraffe individuals. As we explore further PCs, they reveal population structure specific to each taxon (Additional file 2: Fig. S2).

In the admixture analysis (Fig. 1c and Additional file 2: Fig. S3), the plateauing of run likelihoods (Fig. 1d) and the residual fit of the admixture models (Fig. 1e and Additional file 2: Fig. S4) suggest that the number of *K* that better reflect the uppermost level of population structure in the data is $K = 3$. These three ancestry clusters correspond to the focal taxa of the study. As we increase *K*, the population structure within each taxon is revealed and we observe improvements in model fit up to $K = 9$ (Fig. 1c, Fig. 1e, and Additional file 2: Fig. S3 and Fig. S4). This indicates that the admixture model assuming $K = 9$ is the one that best explains the population structure in the data. In this model, most ancestry clusters correspond to groups of geographically close sampling localities. A notable exception is the cluster formed by Nubian giraffe from Gambella National Park (NP), in Ethiopia, and Murchison Falls NP, in Uganda – two locations which are geographically far apart. Individuals from these populations are grouped separately from each other in the PCA (Fig. 1b and Additional file 2: Fig. S2) and the residual fit of the admixture model for $K = 9$ shows a negative correlation between them (Fig. 1e, and Additional file 2: Fig. S4), suggesting that they have different population histories.

We detected signs of admixture from Nubian giraffe in 13 individuals (36.1%) of the reticulated giraffe between $K = 3$–5, with ancestry proportions at $K = 3$ ranging from 0.108–0.434. Like the observations in the PCA, these admixed individuals were all sampled in the Laikipia Plateau. At $K \geq 6$, however, they are assigned to their own cluster with only two

individuals from Loisaba Conservancy (GF292 and GF295) still showing signs of admixture from Nubian giraffe. In the Nubian giraffe, six individuals (19.4%) show admixture from reticulated giraffe between *K* = 3–4. Three of those individuals are from Gambella NP, with ancestry proportions at *K* = 3 ranging from 0.094–0.108, and three are from Murchison Falls NP, with ancestry proportions ranging from 0.002–0.038. However, at *K* = 5 these individuals form a separate cluster which seems to be the source of admixture of the 13 admixed reticulated giraffe individuals. Three individuals (6%) of Masai giraffe s. str. from Tsavo East NP also showed minimal admixture from reticulated giraffe at *K* = 3, with ancestry proportions from 0.023–0.035. However, at higher *K* values these proportions decrease approaching zero.

Nuclear and mitochondrial phylogenies. We reconstructed maximum likelihood trees for two independent datasets: a set of 364,675 genome-wide SNPs from 125 giraffe, and a partitioned alignment of 13 mitochondrial protein-coding genes from 146 giraffe. For taxonomic completeness, both datasets included representatives of all four species and seven subspecies of giraffe with the okapi (*Okapia johnstoni*) as an outgroup. The tree topologies recovered (Fig. 2) are consistent with those reported in previous studies [14, 16, 17]. In the nuclear tree (Fig. 2a), individuals formed reciprocally monophyletic clades corresponding to their respective species with high support (UFboot ≥ 95 and SH-aLTR ≥ 80). Reciprocal monophyly of subspecies, however, was only supported for West African, Kordofan, and Nubian giraffe. Nubian and reticulated giraffe individuals that exhibited admixture signs in the ancestry clustering analysis are placed more externally within the clade of their respective taxa. In the mitochondrial tree (Fig. 2b), Luangwa and Masai giraffe s. str. cannot be distinguished, and the reticulated giraffe is paraphyletic. The grouping of Masai s. str. and South African giraffe is consistent with ancient mitochondrial introgression from Masai s. str. to South African giraffe, as reported in [15, 17], potentially representing a case of mitochondrial capture (i.e., complete replacement of the mitochondrial DNA of one species or population by another). Likewise, the observation of Masai giraffe s. str.

155 individuals carrying reticulated giraffe mitochondrial haplotypes may indicate mitochondrial
156 introgression from reticulated to Masai giraffe s. str., as suggested in [11], although
157 incomplete lineage sorting (ILS) is also a plausible explanation.

158 The individual GF292 carries a Nubian giraffe mitochondrion and falls between the
159 northern giraffe (i.e., West African, Kordofan, and Nubian) and the reticulated giraffe clades
160 in the nuclear phylogeny. That conforms with its high ancestry proportion from Nubian giraffe
161 (0.434 at K = 3 and 0.299 at K = 9; Fig. 1c) and suggests that GF292 is either a recent
162 reticulated × Nubian giraffe hybrid or more likely a backcross from a Nubian giraffe mother.

163 **Migration events and introgression.** We estimated admixture graphs with migration events
164 for Nubian, reticulated, and Masai giraffe s. str. populations (i.e., defined as the resulting
165 clusters at K = 9) using the same dataset used for the SNP-based phylogenomic inference
166 (Fig. 3a). Representatives of all four species and seven subspecies of giraffe were included
167 for taxonomic completeness and the okapi was used as an outgroup. The topology of the
168 estimated admixture graph is consistent with the SNP-based phylogeny. Further, an
169 assessment of the optimal number of migration edges (m) allowed in the graph shows that
170 one migration event (m = 1) is sufficient to explain over 99.8% of the variance in the data
171 (Additional file 2: Fig. S5). However, including a second migration event (m = 2) improves
172 the residual fit of the model (Additional file 2: Fig. S6). At m = 1, a migration event is
173 modelled from Nubian giraffe to the reticulated giraffe from the Laikipia Plateau (locations 8–
174 13), while at m = 2, another migration is modelled from the branch leading to the reticulated
175 giraffe clade to the base of the Masai giraffe s. l., albeit with a lower weight.

176 The inferred admixture graph topology was used as a guide tree to calculate the f-
177 branch (f_b) statistic [33] based on genotype probabilities from the same SNP dataset. That
178 was done for all possible giraffe population trios using the okapi as an outgroup. The f_b
179 assigns evidence for introgression (i.e., f_4-ratio [34] scores) to specific branches on a
180 population/species tree, including internal branches, thus conveying information about the
181 timing of introgression. In our analysis, the f_b identifies a total of 48 signals of excess allele

sharing between the population/species P3 (Fig. 3b, x-axis) and the branch *b* (Fig. 3b, y-axis); however, $f_b \geq 0.05$ in only 17 of them. In particular, the f_b signals suggest gene flow events between reticulated giraffe from Laikipia Plateau (locations 8–13) and the branches leading to Kordofan + Nubian giraffe ($f_b = 0.16$) and to Nubian giraffe populations ($f_b = 0.21$). Weaker f_b signals also indicate gene flow between reticulated giraffe populations (locations 8–13 and 14–18) and the branch leading to Masai s. str. + Luangwa giraffe ($f_b = 0.09$ in both cases). However, the strongest identified f_b signal ($f_b = 0.45$) corresponds to gene flow between Masai giraffe s. str. populations from locations 24–30 (i.e., Amboseli NP, Hell's Gate NP, Mbirikani, Nairobi NP, Naivasha, Ngong, and Tsavo West NP) and the branch leading to Masai giraffe s. str. populations from the Selous (location 19) and Masai Mara Game Reserves (locations 20–23). In all those cases, gene flow from P3 (x-axis) into branch *b* (y-axis) also generated horizontal lines of correlated f_b signals between branch *b* and lineages related to P3 due to their shared ancestry with P3 [35].

Contemporary gene flow. We estimated both directionality and rates of contemporary migration (last two generations) between Nubian, reticulated, and Masai giraffe s. str. based on a subset of 8,137 unlinked SNPs from 97 individuals with median read depth ≥ 8. The highest mean posterior migration rate is observed from Nubian to reticulated giraffe, where 2% of the individuals in the reticulated giraffe are estimated to be migrants derived from the Nubian giraffe (per generation) (Fig. 3c and Additional file 3: Table S2). Migration rates inferred between other species in any direction are ≤ 1.1%. However, all 95% credible intervals for migration rates include zero, and therefore absence of recent gene flow cannot be statistically rejected (Additional file 3: Table S2).

Demographic reconstruction. Reconstruction of population size changes over the recent past based on the site frequency spectrum (SFS) reveals a general decrease in effective population sizes (N_e) for the three analyzed giraffe taxa (Fig. 4). We observe similar but unsynchronized demographic trends with an accentuated population bottleneck (Nubian: ~6.5–18 ka ago; reticulated: ~28–54 ka ago, Masai s. str.: ~10.5–25 ka ago) intercalating

209 periods of relative stability at higher N_e (Nubian: ~0.7–5 ka and ~18–76 ka ago; reticulated:
210 ~3–20 ka and ~54–140 ka ago, Masai s. str.: ~2–7 ka and ~25–60 ka ago). The Nubian and
211 the reticulated giraffe experienced a sharp decline between ~0.48–0.7 ka and ~2.5–3 ka
212 ago, respectively, towards an approximately constant N_e, while the Masai giraffe s. str.
213 shows a gradual decline between ~0.7–2 ka before reaching relative stability. Population
214 bottlenecks older than 50 ka are observed for all three giraffe taxa; however, these cannot
215 be interpreted reliably due to limitations of SFS-based demographic methods for ancestral
216 time spans [36]. Overall, median N_e dropped from their highest ancestral estimates of
217 ~62,400 to currently ~2,700 for the Nubian giraffe, from ~130,200 to a present ~5,500 for the
218 reticulated giraffe, and from ~29,500 to ~1,700 for the Masai giraffe s. str.

219

220 **Discussion**

221 The number of giraffe species has been a subject of debate, particularly the question
222 whether northern and reticulated giraffe should be considered separate species [14–17]. As
223 the 'species' is predominantly used as the fundamental unit of conservation and as a metric
224 of biodiversity [37], understanding species distinction is key for accurate conservation status
225 assessments that can effectively guide conservation efforts [18]. In addition, failing to identify
226 or neglecting admixed populations and hybrids can be detrimental to conservation policy
227 making, and thus to species conservation [38]. Our findings corroborate significant genetic
228 divergence between northern, reticulated, and Masai giraffe s. l., as shown by the distinct
229 PCA and admixture clusters, as well as their reciprocal monophyly in the SNP phylogeny. In
230 the mitochondrial tree, the nesting of northern giraffe within a paraphyletic reticulated giraffe
231 may be explained by ILS resulting from peripatric or "budding" speciation [39] in peripheral
232 populations of reticulated giraffe. However, under peripatric speciation, parallel patterns of
233 paraphyly are expected across nuclear and mitochondrial loci [39], and this was not
234 observed in our dataset. Furthermore, while three Masai giraffe s. str. individuals were

identified carrying reticulated giraffe mitochondria, we cannot confidently distinguish between ancient mitochondrial introgression and/or ILS as the likely cause.

As demonstrated, while introgressive hybridization between Masai giraffe s. str. and other giraffe taxa in Kenya is minimal, admixture between Nubian – the easternmost subspecies of the northern giraffe – and reticulated giraffe seems to be asymmetrical towards the latter and restricted to a contact zone in the Laikipia Plateau. Although hybridization among giraffe in that region has been previously conjectured, with the observation of individuals exhibiting intermediate phenotypes (i.e., pelage pattern) [26, 28], we provide the first genomic evidence of its occurrence. This finding reinforces the unreliability of morphological characters such as pelage pattern for the identification of giraffe (sub)species, especially at a local scale [27]. Moreover, while the estimated migration events in the admixture graph and the f_b statistic support gene flow between Nubian and reticulated giraffe across the Laikipia Plateau, they also suggest that such an event is most likely ancestral. Consistent with that, estimates of contemporary migration rates were low and not statistically significant suggesting that current gene flow is limited. In fact, only two reticulated giraffe individuals consistently show signs of admixture from Nubian giraffe upon deeper investigation of population structure at $K \geq 6$. Habitat loss and fragmentation, linked with human population growth, has drastically reduced the natural range of the Nubian giraffe in Kenya, such that most of its present-day populations are a result of extralimital translocations from a source population near Eldoret during the 1970s [18, 40]. All these introductions were into government or private fenced wildlife areas, which would have further restricted gene flow between Nubian and reticulated giraffe over the last few generations. More importantly, however, the overall genomic integrity of the parent taxa despite the existence of a contact zone suggests reproductive isolation and can be interpreted as support for species status [41].

Previous studies hypothesized that the divergence between giraffe lineages in East Africa could be linked to climatic oscillations associated with Earth's precession cycles

during the Pleistocene [11, 24]. Divergence could have been triggered either by spatially and temporally contrasting rainfall regimes that persisted until the Early Holocene, or by repeated expansion and contraction of the savannah habitat resulting in periodic isolation in refugia [24]. Nevertheless, the long-term maintenance of genetic distinctiveness between them appears to correlate with regionally distinct rainfall patterns in East Africa and the associated differences in timing of emergence and availability of browse [24]. Reproductive asynchrony resulting from potential local adaptation of the reproductive cycle to the differential timing of green-up may explain such correlation [11, 24]. That would imply that hybrids may display reduced fitness if they were born in an unfavorable season [11, 24], and result in negative selection against them, which would likely restrain introgression to contact zones [41]. Introgressive hybridization from Nubian into reticulated giraffe in Laikipia may have occurred under these conditions. Nubian giraffe populations have been shown to exhibit temporal and seasonal migration patterns [42]. In our study, this is particularly important considering the proximity and lack of a geographic barrier between the Lake Baringo Basin, a historical Nubian giraffe stronghold, and the Laikipia Plateau, which currently holds ~28% of the extant reticulated giraffe population. Conversely, the absence of substantial admixture involving Masai giraffe s. str. could reflect stronger selection against its hybrids. Seasonal variation in habitat use (i.e., resource tracking) and pelage-based assortative mating could also affect the maintenance of genetic divergence and broad-scale phenotypic differences between the taxa [11, 24].

Our reconstruction of demographic changes for the three focal taxa in the recent past expands previous inferences made for the distant past [17], and provide a more complete picture of their population history. In general, estimates of N_e for the three giraffe lineages were higher during the Late Pleistocene than they are today. Population reductions observed since the Late Pleistocene–Holocene transition could be linked to a period of wetter conditions and associated contraction of savannahs that lasted from ~5.5–14.8 ka ago [43], and later to the spread of pastoralism in East Africa from ~4.7 ka ago onwards [44]. N_e

values estimated at the present are reasonable, and expectedly lower [45] than the current estimated census population sizes (N_c) [18]. The reticulated giraffe has the highest present-day N_e (~5,500) among the three taxa, while the Masai giraffe s. str. has the lowest (~1,700). Furthermore, the Masai giraffe s. str. shows the largest difference between present-day N_e (~1,700) and N_c (~44,700 [18]), followed by the reticulated giraffe (N_e = ~5,500 and N_c = ~16,000 [18]), and the Nubian giraffe (N_e = ~2,700 and N_c = ~3,000 [18]). These observations are in line with previous findings of genomic diversity that is higher in reticulated, moderate in Nubian, and lower in Masai giraffe s. str. [17].

Conclusions

Our findings reinforce the classification of giraffe into the four species (i.e., northern, reticulated, Masai s. l., and southern giraffe) proposed in [13, 14, 17], by clearly separating northern from reticulated giraffe, with limited recent introgression reflecting effective reproductive isolation. These results have valuable and direct conservation implications for the management of giraffe in Kenya and more broadly throughout their range in Africa. As the three species present in Kenya are genetically distinct, it is important that future conservation interventions, such as translocations, take these findings into account to avoid mixing distinct (sub)species, hence maintaining their unique biodiversity [46]. The outcome of this study is critical to appropriate re-classification of giraffe on the IUCN Red List and in turn informing targeted conservation actions for each taxon, particularly for African range states and international convention reviews such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) [47, 48]. Moreover, the comprehensive genomic dataset made available here constitutes an important resource for future studies of local adaptation in the giraffe lineages in East Africa. By identifying loci under selection and deeply characterizing the genetic composition of admixed/hybrid individuals, it might be possible to shed light on the potential effects of such loci on the fitness of giraffe hybrids in nature.

Methods

Sampling. Skin biopsy samples from 113 wild giraffe (Nubian, *n* = 32; reticulated, *n* = 37; Masai s. str., *n* = 44) from different parts of Kenya were collected as a collaboration between the Giraffe Conservation Foundation (GCF) and Kenya Wildlife Service (KWS), together with local partners, via remote biopsy darting and carcasses, and preserved in absolute ethanol. Sampling was conducted with the appropriate access and research permits from the Kenyan authorities. Sampling locations and sample details are presented in Fig. 1a and Additional file 1: Table S1. Short reads from wild individuals of West African (*n* = 5), Kordofan (*n* = 5), Nubian (*n* = 6), reticulated (*n* = 3), Masai s. str. (*n* = 6) , Luangwa (*n* = 6), South African (*n* = 6), and Angolan giraffe (*n* = 6) analyzed in Coimbra et al. [17] and Agaba et al. [49] were added to the new dataset for a comprehensive representation of all four species and seven subspecies of giraffe (Additional file 1: Table S1). The okapi from Agaba et al. [49] was included in analyses that required an outgroup.

Whole-genome sequencing. DNA was extracted using either a NucleoSpin Tissue Kit (Macherey-Nagel) or the phenol-chloroform protocol [50]. Sequencing libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Inc.) at Novogene and sequenced on an Illumina NovaSeq 6000 (2 × 150 bp, 350 bp insert size).

Read processing. Quality control of short reads was done in fastp v0.20.0 [51] with base correction and low complexity filter enabled. Adapters and polyG stretches in read tails were automatically detected and removed. Trimming was performed in a 4-bp sliding window (option --cut_right for reads from Agaba et al. [49] and --cut_tail for the remaining) with a required mean base quality ≥ 15. Reads shorter than 36 bp, composed of > 40% of bases with quality < 15, or containing > 5 Ns were discarded.

Read mapping. Reads were mapped against a chromosome-level Masai giraffe s. str. genome assembly [32] (GenBank: GCA_013496395) with BWA-MEM v0.7.17-r1188 [52].

The resulting SAM files were coordinate-sorted, converted to BAM, and merged to sample-level with samtools v1.10 [53]. Duplicate reads in the BAM files were marked with MarkDuplicates from Picard v2.21.7 (http://broadinstitute.github.io/picard/) and regions around indels were realigned with GATK v3.8.1 [54]. Reads mapped to repetitive regions identified with RepeatMasker v4.0.7-open [55] and to sex chromosomes, unmapped reads, and reads flagged with bits ≥ 256 were removed from the BAM files with samtools. Only reads mapped in a proper pair were retained.

SNP calling and linkage pruning. SNPs were called per species in Nubian, reticulated, and Masai giraffe s. str. individuals with median read depth ≥ 2 (Additional file 1: Table S1). Genotype likelihoods were estimated in ANGSD v0.933 [56] with options -GL 1 -baq 2. Minimum mapping and base quality scores were set to 30 and depth thresholds were set to $d \pm (5 \times MAD)$, where d is the median and MAD is the median absolute deviation of the global site depth distribution. Only biallelic SNPs called with a p-value $< 1 \times 10^{-6}$, present in at least 90% of the individuals, and with a minor allele frequency (MAF) of at least 5% were considered. SNPs were tested for strand bias, heterozygous bias, and deviation from Hardy-Weinberg equilibrium (HWE) and discarded if any of the resulting p-values was below 1×10^{-6}.

Each species' SNP set was then independently pruned for linkage disequilibrium (LD) with ngsLD v1.1.1 [57]. We estimated r^2 values for SNP pairs up to 500 kbp apart as a proxy for LD and plotted linkage decay curves per species from a random sample of 0.05% of the estimated r^2 values (Additional file 2: Fig. S7). Appropriate thresholds for linkage pruning were selected based on each species' linkage decay curve. SNPs were pruned assuming a maximum pairwise distance of 100 kbp for all species and a minimum r^2 of 0.10 for reticulated and Masai giraffe and 0.15 for the Nubian giraffe. The resulting pruned SNPs from each species were concatenated and used as input in a second SNP calling round to generate a combined dataset with individuals from the three species. SNPs were jointly

called in ANGSD with the -sites option and no MAF, HWE, or SNP *p*-value filters were used. All remaining settings were as described above.

Relatedness. The combined SNP dataset generated above was used to estimate pairwise relatedness in NGSremix v1.0.0 [58], which accounts for individuals with admixed ancestry. NGSremix additionally requires admixture proportions and ancestral allele frequencies as inputs, which were obtained from the run with the highest log-likelihood out of 100 runs of NGSadmix v32 [59] assuming three ancestry components (*K*). A custom R script was used to identify and select closely related individuals that, when removed from the dataset, would maximize the reduction of the overall relatedness in the data while minimizing sample loss (Additional file 1: Table S1 and Additional file 2: Fig. S1). These individuals were removed from all further analyses. A final round of joint SNP calling with ANGSD was performed as described above to obtain a combined SNP dataset for Nubian, reticulated, and Masai giraffe s. str. that was LD pruned and filtered against relatedness.

Population structure and admixture. Genotype likelihoods of unlinked SNPs estimated in ANGSD were used to calculate a covariance matrix in PCAngsd v1.03 [60]. A PCA was then performed using the prcomp() function in R v4.2.2 [61]. Ancestry clusters and individual ancestry proportions were inferred in NGSadmix assuming a *K* value ranging from 1 to 11. The analysis was repeated 100 times per *K* with different random seeds and the replicate with the highest log-likelihood for each $K \geq 2$ was shown as an admixture plot. The fit of the resulting admixture models was assessed based on the pairwise correlation of residuals between individuals estimated with evalAdmix v0.962 [62].

SNP-based phylogenomic inference. Genotypes were jointly called in individuals from all giraffe species and subspecies with median read depth ≥ 8 with the okapi as an outgroup (Additional file 1: Table S1). Genotype calling was performed using bcftools v1.17 [53] mpileup + call pipeline, with option --full-BAQ and minimum mapping and base quality set to 30. Samples were grouped per (sub)species (--group-samples) and HWE assumption was applied within but not across groups. The commands filter and view were used to convert

genotypes with GQ < 20 to missing data and filter for biallelic SNPs with a fraction of missing genotypes ≤ 0.1, QUAL ≥ 30, MQ ≥ 30, and within depth thresholds calculated as described for ANGSD. To reduce the impact of LD, we randomly sampled ~1% (462,697) of all SNPs using vcfrandomsample from vcflib v1.0.3 [63]. The called genotypes from the subsampled VCF were used create a matrix for phylogenetic analysis in PHYLIP format with vcf2phylip v2.8 [64]. After removing constant, partially constant, and ambiguously constant sites from the SNP matrix, a maximum likelihood phylogeny was constructed in IQ-TREE v2.2.2.3 [65] based on 364,675 SNPs. Ultrafast model selection [66] between nucleotide substitution models with ascertainment bias correction [67] was enabled. Branch support was assessed by 1,000 replicates of the ultrafast bootstrap approximation (UFBoot) [68], with hill-climbing nearest neighbor interchange (NNI) optimization, and the Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) [69]. The tree was plotted with ggtree v3.6.2 [70].

Assembly and phylogeny of mitochondrial genomes. Mitogenomes were assembled *de novo* from the unprocessed reads using GetOrganelle v1.7.4 [71] with options –-fast -k 21,45,65,85,105 -F animal_mt. In some instances, fine-tuning parameters -w and –max-n-words was necessary to obtain a complete circular genome. Mitogenome assembly sequences were visually inspected and curated (i.e., corrected directionality, circularized, adjusted starting nucleotide) on Geneious Prime v2020.1.2 (https://www.geneious.com/).

Sequences of all 13 mitochondrial protein-coding genes were extracted from the assemblies and aligned to sequences of wild giraffe analyzed in Coimbra et al. [17] and Hassanin et al. [72] (GenBank: JN632645) with the L-INS-i algorithm in MAFFT v7.475 [73]. The okapi (GenBank: JN632674) [72] was also included as an outgroup for phylogenetic inference. A maximum likelihood phylogeny was constructed in IQ-TREE through a partitioned analysis [74] of the protein-coding gene alignments. Ultrafast model selection between codon models was enabled assuming the vertebrate mitochondrial genetic code.

Branch support was assessed with 1,000 replicates of the UFBoot and the SH-aLRT. The tree was plotted with ggtree.

Inference of migration events. The topology and migration events between the Nubian, reticulated, and Masai giraffe s. str. populations (defined as the clusters resulting from the best fitting admixture model) were inferred as admixture graphs with TreeMix v1.13-r231 [75] and OrientAGraph v1.0 [76]. The subsampled VCF generated for the SNP-based phylogenomic inference was processed with PLINK v1.9 [77] and plink2treemix.py (https://bit.ly/3LCcNW4) to obtain allele counts per population as input. TreeMix was ran using blocks of 100 SNPs and assuming the number of migration edges (*m*) ranging from 0 to 5 for 50 independent optimization runs per *m*. The okapi was used to root the graph. A range of *m* values to be further explored was determined by looking at the Δm and the percentage of explained variance estimated with OptM v0.1.6 [78]. We then ran OrientAGraph with options -mlno -allmigs for 10 bootstrap replicates assuming the selected *m* values. OrientAGraph improves TreeMix's heuristics with an exhaustive search for a maximum likelihood network orientation (MLNO) resulting in graphs with higher likelihood scores and topological accuracy. The run with the highest log-likelihood per *m* value was selected.

Test for introgression. Genotype probabilities from the SNP dataset used to infer migration events were used to calculate Patterson's *D*, f_4-ratio [34], and *f*-branch (f_b) [33] statistics for all possible giraffe population trios using Dsuite v0.5-52 [35] with the okapi as an outgroup. The f_b is of particular interest as it can disentangle correlated f_4-ratio estimates and assign evidence for introgression to specific, possibly internal, branches on a phylogeny given that they can be tested under a ((P1, P2) P3, Outgroup) topology. The admixture graph topology reconstructed by OrientAGraph, which was identical for $m = 1$ and $m = 2$, was used as the guide tree for the f_b estimation. Statistical significance was assessed through block-jackknifing with 100 equally sized blocks.

Contemporary migration rates. Directionality and rates of contemporary migration between Nubian, reticulated, and Masai s. str. giraffe were estimated with BA3-SNPs v1.1 [79, 80]. First, a VCF file was generated by jointly calling genotypes in all individuals with median read depth ≥ 8 (Additional file 1: Table S1). Unlinked SNP sites identified in the second round of ANGSD were supplied to bcftools' mpileup + call pipeline. Genotypes were called and sites were filtered as described for the SNP-based phylogenomic inference. We then randomly sampled ~2% (8,137) of all SNPs using vcflib's vcfrandomsample and converted the resulting VCF to the input format for BA3-SNPs with Stacks v2.59 [81] and stacksStr2immanc.pl (https://bit.ly/34fJdUz). Mixing parameters for BA3-SNPs were automatically tuned to achieve acceptance rates between 0.2 and 0.6 with BA3-SNPs-autotune v2.1.2 [82] by conducting short exploratory runs of 2.5 million iterations with a burn-in phase of 500,000 steps. The final BA3-SNPs run used 22 million iterations, a burn-in phase of 2 million steps, a sampling interval of 2,000 iterations, and the tuned mixing parameters (-m 0.1000 -a 0.2125 -f 0.0125). To assess chain convergence, the analysis was repeated three times, each starting from a different random seed, and the log probabilities of each run were plotted (Additional file 2: Fig. S8). The run with the smallest Bayesian deviance was selected [83] and the estimated migration rates were shown as a circular plot. We also constructed 95% credible sets using the formula $m \pm (1.96 \times sdev)$, where m is the posterior mean migration rate and *sdev* is the standard deviation of the marginal posterior distribution [79]. Migration rates were considered significant if the credible set did not include zero.

Demographic reconstruction. Recent changes in N_e of Nubian, reticulated, and Masai giraffe s. str. were assessed based on the SFS. First, a genome consensus sequence was generated for the okapi using ANGSD with option -doFasta 1 to polarize SNPs during the SFS estimation. We enabled -baq 2 and discarded sites with mapping or base qualities < 30 or depth below 4 or above the 95th percentile of the sample's depth distribution. We then calculated site allele frequencies per species in ANGSD with option -doSaf 1 and the okapi

consensus sequence as ancestral. Individuals with median read depth < 2 were not included and quality filters were set as described for SNP calling. No HWE, MAF, and SNP *p*-value filters were used [84]. Site allele frequencies were converted into the unfolded SFS with ANGSD's realSFS. Demographic histories were reconstructed from the unfolded SFS using Stairway Plot v2.1.1 [85], after masking singletons, with a mutation rate of 2.12×10^{-8} substitutions per site per generation estimated for the giraffe [86] and a generation length of 10 years [87].

Supplementary Information

The online version contains supplementary material available at XXX.

Additional file 1: Table S1. Sample details and mapping statistics for analyzed individuals.

Additional file 2: Figures S1–S8. Fig. S1. Relatedness between pairs of individuals. **Fig. S2.** Principal component analysis (PCA). **Fig. S3.** Admixture analyses assuming a varying number of ancestry components (*K*). **Fig. S4.** Assessment of the fit of admixture models assuming K = 1–11 based on the correlation of residuals. **Fig. S5.** OptM output for the TreeMix runs with migration edges (*m*) ranging from 0 to 5. **Fig. S6.** Admixture graphs of giraffe populations and their corresponding residual fit. **Fig. S7.** Linkage disequilibrium (LD) decay in Nubian, reticulated, and Masai giraffe s. str. **Fig. S8.** Log probability trace and Bayesian deviance (D) for each BA3-SNPs run.

Additional file 3: Table S2. Posterior mean migration rates among Nubian, reticulated, and Masai giraffe s. str.

Declarations

Ethics approval and consent to participate: Sampling of giraffe skin biopsies was conducted under the appropriate access and research permits from the Kenyan authorities (#NEMA/AGR/109/2018/93, #KWS/BRM/5001, and #NACOSTI/P/18/50967/20704).

Consent for publication: Not applicable.

Availability of data and materials: Raw sequencing reads generated in this study are available at NCBI Short Read Archive under the BioProject accession PRJNA815626. Nucleotide sequences of mitochondrial genomes assembled in this study are available at GenBank under the accession numbers OM973995–OM974107. The code used to process and analyze the data generated in this study is available at GitHub (https://bit.ly/3LmwG4p).

Competing interests: The authors declare that they have no competing interests.

Funding: The present study is a product of the Centre for Translational Biodiversity Genomics (LOEWE-TBG) as part of the “LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz” program of Hesse’s Ministry of Higher Education, Research, and the Arts as well as the Leibniz Association. Field sampling for the study was provided by the Giraffe Conservation Foundation.

Authors' contributions: Conceptualization, RTFC, SW, JF and AJ; methodology, RTFC; software, RTFC; validation, RTFC; formal analysis, RTFC; investigation, RTFC; resources, AM, SF, MO, DM, SM, JS-D, JF and AJ; data curation, RTFC; writing—original draft, RTFC, SW, JF and AJ; writing—review and editing, RTFC, SW, AM, SF, MO, DM, SM, JS-D, JF and AJ; visualization, RTFC; supervision, AJ; project administration, AJ; funding acquisition, SF, JF and AJ.

Acknowledgements: We thank an array of partners, in particular government and NGO partners across Kenya who collaborated with and/or financially supported the Giraffe Conservation Foundation to permit, collect and include samples in this analysis, including Cleveland Metroparks Zoo, Governments of Botswana, Chad, Ethiopia, Kenya, Namibia, Niger, Tanzania, Uganda, and Zambia, Ivan Carter Wildlife Conservation Alliance, and San

Diego Zoo Wildlife Alliance. We also thank Emma Vinson for her assistance in coding the R script used for relatedness filtering.

References

1. Stankowski S, Ravinet M. Defining the speciation continuum. Evolution. 2021;75:1256–73.

2. Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. Annu Rev Genet. 2021;55:265–83.

3. Vonlanthen P, Bittner D, Hudson AG, Young KA, Müller R, Lundsgaard-Hansen B, et al. Eutrophication causes speciation reversal in whitefish adaptive radiations. Nature. 2012;482:357–62.

4. Rieseberg LH, Van Fossen C, Desrochers AM. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature. 1995;375:313–6.

5. Mallet J, Besansky N, Hahn MW. How reticulated are species? BioEssays. 2016;38:140–9.

6. Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. J Hered. 2014;105:795–809.

7. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20:229–37.

8. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 2013;23:1817–28.

9. Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518:371–5.

546 10. Garcia-Erill G, Kjær MM, Albrechtsen A, Siegismund HR, Heller R. Vicariance followed
547 by secondary gene flow in a young gazelle species complex. Mol Ecol. 2021;30:528–44.

548 11. Brown DM, Brenneman RA, Koepfli K-P, Pollinger JP, Milá B, Georgiadis NJ, et al.
549 Extensive population genetic structure in the giraffe. BMC Biol. 2007;5:57.

550 12. Groves C, Grubb P. Giraffidae. In: Ungulate taxonomy. Baltimore, MD: Johns Hopkins
551 Univ. Press; 2011. p. 64–70.

552 13. Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, et al. Multi-locus
553 analyses reveal four giraffe species instead of one. Curr Biol. 2016;26:2543–9.

554 14. Winter S, Fennessy J, Janke A. Limited introgression supports division of giraffe into four
555 species. Ecol Evol. 2018;8:10156–65.

556 15. Petzold A, Hassanin A. A comparative approach for species delimitation based on
557 multiple methods of multi-locus DNA sequence analysis: a case study of the genus *Giraffa*
558 (Mammalia, Cetartiodactyla). PLoS One. 2020;15:e0217956.

559 16. Petzold A, Magnant A-S, Edderai D, Chardonnet B, Rigoulet J, Saint-Jalme M, et al. First
560 insights into past biodiversity of giraffes based on mitochondrial sequences from museum
561 specimens. Eur J Taxon. 2020. https://doi.org/10.5852/ejt.2020.703.

562 17. Coimbra RTF, Winter S, Kumar V, Koepfli K-P, Gooley RM, Dobrynin P, et al. Whole-
563 genome analysis of giraffe supports four distinct species. Curr Biol. 2021;31:2929-2938.e5.

564 18. Brown MB, Kulkarni T, Ferguson S, Fennessy S, Muneza A, Stabach JA, et al.
565 Conservation status of giraffe: evaluating contemporary distribution and abundance with
566 evolving taxonomic perspectives. In: DellaSala DA, Goldstein MI, editors. Imperiled: the
567 encyclopedia of conservation. Oxford: Elsevier; 2022. p. 471–87.

568 19. Le Pendu Y, Ciofolo I. Seasonal movements of giraffes in Niger. J Trop Ecol.
569 1999;15:341–53.

570 20. Fennessy J. Home range and seasonal movements of *Giraffa camelopardalis angolensis*
571 in the northern Namib Desert. Afr J Ecol. 2009;47:318–27.

572 21. Dagg AI. External features of giraffe. Mammalia. 1968;32:657–69.

573 22. Gray AP. Mammalian hybrids: a check-list with bibliography. 2nd edition. Farnham
574 Royal, UK: Commonwealth Agricultural Bureaux; 1972.

575 23. Lackey LB. Giraffe, Giraffa camelopardalis, North American regional/global studbook. 8th
576 edition. 2011.

577 24. Thomassen HA, Freedman AH, Brown DM, Buermann W, Jacobs DK. Regional
578 differences in seasonal timing of rainfall discriminate between genetically distinct East
579 African giraffe taxa. PLoS One. 2013;8:e77191.

580 25. Lydekker R. Two undescribed giraffes. Nature. 1911;87:484.

581 26. Stott K. Giraffe intergradation in Kenya. J Mammal. 1959;40:251.

582 27. Dagg AI. The subspeciation of the giraffe. J Mammal. 1962;43:550–2.

583 28. Stott KW, Selsor CJ. Further remarks on giraffe intergradation in Kenya and unreported
584 marking variations in reticulated and Masai giraffes. Mammalia. 1981;45:261–3.

585 29. Gompert Z, Mandeville EG, Buerkle CA. Analysis of population genomic data from hybrid
586 zones. Annu Rev Ecol Evol Syst. 2017;48:207–29.

587 30. Feder JL, Egan SP, Nosil Patrik. The genomics of speciation-with-gene-flow. Trends
588 Genet. 2012;28:342–50.

589 31. Quilodrán CS, Montoya-Burgos JI, Currat M. Harmonizing hybridization dissonance in
590 conservation. Commun Biol. 2020;3:391.

591 32. Farré M, Li Q, Darolti I, Zhou Y, Damas J, Proskuryakova AA, et al. An integrated
592 chromosome-scale genome assembly of the Masai giraffe (*Giraffa camelopardalis*
593 *tippelskirchi*). Gigascience. 2019;8.

33. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol. 2018;2:1940–55.

34. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. Genetics. 2012;192:1065–93.

35. Malinsky M, Matschiner M, Svardal H. Dsuite: fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021;21:584–95.

36. Liu X, Fu Y-X. Exploring population size changes using SNP frequency spectra. Nat Genet. 2015;47:555–9.

37. Coates DJ, Byrne M, Moritz C. Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. Front Ecol Evol. 2018;6:165.

38. Bauer H, Tehou AC, Gueye M, Garba H, Doamba B, Diouck D, et al. Ignoring species hybrids in the IUCN Red List assessments for African elephants may bias conservation policy. Nat Ecol Evol. 2021;5:1050–1.

39. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 2003;34:397–423.

40. Brenneman RA, Bagine RK, Brown DM, Ndetei Robert, Louis Jr. EE. Implications of closed ecosystem conservation management: the decline of Rothschild's giraffe (*Giraffa camelopardalis rothschildi*) in Lake Nakuru National Park, Kenya. Afr J Ecol. 2009;47:711–9.

41. Wielstra B. Hybrid zones. Curr Biol. 2021;31:R108–9.

42. Brown MB, Bolger DT. Male-biased partial migration in a giraffe population. Front Ecol Evol. 2020;7.

617 43. Shanahan TM, McKay NP, Hughen KA, Overpeck JT, Otto-Bliesner B, Heil CW, et al.
618 The time-transgressive termination of the African Humid Period. Nat Geosci. 2015;8:140–4.

619 44. Chritz KL, Cerling TE, Freeman KH, Hildebrand EA, Janzen A, Prendergast ME. Climate,
620 ecology, and the spread of herding in eastern Africa. Quat Sci Rev. 2019;204:119–32.

621 45. Frankham R. Effective population size/adult population size ratios in wildlife: a review.
622 Genet Res. 1995;66:95–107.

623 46. Fennessy J, Bower V, Castles M, Fennessy S, Brown M, Hoffman R, et al. A journey of
624 giraffe – a practical guide to wild giraffe translocations. Giraffe Conservation Foundation.
625 2022. https://library.giraffeconservation.org/download/5814/. Accessed 30 May 2023.

626 47. Kenya Wildlife Service. National recovery and action plan for giraffe (*Giraffa*
627 *camelopardalis*) in Kenya (2018-2022). Kenya Wildlife Service. 2018.
628 https://giraffeconservation.org/wp-content/uploads/2019/10/National-Recovery-and-Action-
629 Plan-for-Giraffe-in-Kenya-2018-2022.pdf. Accessed 30 May 2023.

630 48. Convention on International Trade in Endangered Species of Wild Fauna and Flora.
631 Proposals for amendment of Appendices I and II. Eighteenth meeting of the Conference of
632 the Parties, Geneva (Switzerland), 17-28 August 2019. 2019.
633 https://cites.org/eng/cop/18/proposals_for_amendment. Accessed 30 May 2023.

634 49. Agaba M, Ishengoma E, Miller WC, McGrath BC, Hudson CN, Bedoya Reina OC, et al.
635 Giraffe genome sequence reveals clues to its unique morphology and physiology. Nat
636 Commun. 2016;7:11519.

637 50. Sambrook J, Russel DW. Molecular cloning: a laboratory manual. 3rd edition. Cold
638 Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.

639 51. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
640 Bioinformatics. 2018;34:i884–90.

641 52. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
642 ArXiv. 2013. https://doi.org/10.48550/arXiv.1303.3997.

643 53. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
644 SAMtools and BCFtools. Gigascience. 2021;10:giab008.

645 54. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
646 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
647 sequencing data. Genome Res. 2010;20:1297–303.

648 55. Smit AFA, Hubley R, Green P. RepeatMasker open-4.0. 2015. http://repeatmasker.org.
649 Accessed 12 Apr 2019.

650 56. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation
651 sequencing data. BMC Bioinformatics. 2014;15:356.

652 57. Fox EA, Wright AE, Fumagalli M, Vieira FG. ngsLD: evaluating linkage disequilibrium
653 using genotype likelihoods. Bioinformatics. 2019;35:3855–6.

654 58. Nøhr AK, Hanghøj K, Garcia-Erill G, Li Z, Moltke I, Albrechtsen A. NGSremix: a software
655 tool for estimating pairwise relatedness between admixed individuals from next-generation
656 sequencing data. G3 Genes|Genomes|Genetics. 2021;11:jkab174.

657 59. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions
658 from next generation sequencing data. Genetics. 2013;195:693–702.

659 60. Meisner J, Albrechtsen A. Inferring population structure and admixture proportions in
660 low-depth NGS data. Genetics. 2018;210:719–31.

661 61. R Core Team. R: a language and environment for statistical computing. 2022.
662 https://www.r-project.org/. Accessed 31 Oct 2022.

663 62. Garcia-Erill G, Albrechtsen A. Evaluation of model fit of inferred admixture proportions.
664 Mol Ecol Resour. 2020;20:936–49.

665 63. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free
666 software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and
667 slivar. PLoS Comput Biol. 2022;18:e1009123.

668 64. Ortiz EM. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for
669 phylogenetic analysis. Zenodo. 2019. https://doi.org/10.5281/zenodo.2540861. Accessed 1
670 Jun 2023.

671 65. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et
672 al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic
673 era. Mol Biol Evol. 2020;37:1530–4.

674 66. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
675 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

676 67. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological
677 character data. Syst Biol. 2001;50:913–25.

678 68. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the
679 ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.

680 69. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms
681 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
682 PhyML 3.0. Syst Biol. 2010;59:307–21.

683 70. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and
684 annotation of phylogenetic trees with their covariates and other associated data. Methods
685 Ecol Evol. 2017;8:28–36.

686 71. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast
687 and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol.
688 2020;21:241.

689 72. Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, et al.
690 Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as
691 revealed by a comprehensive analysis of mitochondrial genomes. C R Biol. 2012;335:32–50.

692 73. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
693 improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

694 74. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic
695 inference from supermatrices. Syst Biol. 2016;65:997–1008.

696 75. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide
697 allele frequency data. PLoS Genet. 2012;8:e1002967.

698 76. Molloy EK, Durvasula A, Sankararaman S. Advancing admixture graph estimation via
699 maximum likelihood network orientation. Bioinformatics. 2021;37 Supplement_1:i142–50.

700 77. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation
701 PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

702 78. Fitak RR. OptM: estimating the optimal number of migration edges on population trees
703 using Treemix. Biol Methods Protoc. 2021;6:bpab017.

704 79. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus
705 genotypes. Genetics. 2003;163:1177–91.

706 80. Mussmann SM, Douglas MR, Chafin TK, Douglas ME. BA3-SNPs: contemporary
707 migration reconfigured in BayesAss for next-generation sequence data. Methods Ecol Evol.
708 2019;10:1808–13.

709 81. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: analytical methods for
710 paired-end sequencing improve RADseq-based population genomics. Mol Ecol.
711 2019;28:4737–54.

712 82. Mussmann S, Chafin T. stevemussmann/BA3-SNPS-autotune: BA3-SNPs-autotune
713 v2.1.2. Zenodo. 2020. https://doi.org/10.5281/zenodo.4017836. Accessed 1 Jun 2023.

714 83. Meirmans PG. Nonconvergence in Bayesian estimation of migration rates. Mol Ecol
715 Resour. 2014;14:726–33.

716 84. Matz M V. Fantastic beasts and how to sequence them: ecological genomics for obscure
717 model organisms. Trends Genet. 2018;34:121–32.

718 85. Liu X, Fu Y-X. Stairway Plot 2: demographic history inference with folded SNP frequency
719 spectra. Genome Biol. 2020;21:280.

720 86. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome
721 sequencing provides insights into their evolution and distinct traits. Science.
722 2019;364:eaav6202.

723 87. Muller Z, Bercovitch F, Brand R, Brown D, Brown M, Bolger D, et al. *Giraffa*
724 *camelopardalis* (amended version of 2016 assessment). The IUCN Red List of Threatened
725 Species. 2018. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T9194A136266699.en.
726 Accessed 1 Jun 2023.

727

Figure Legends

Fig. 1. Population structure of Nubian, reticulated, and Masai giraffe s. str. (**a**) Geographical distribution of Nubian, reticulated and Masai giraffe s. str. (colored shadings) in East Africa and sampling locations (colored shapes and numbers). Hatched areas show estimated range of Nubian and reticulated giraffe populations. (**b**) PCA of 484,876 unlinked SNPs from 116 individuals representing Nubian, reticulated, and Masai giraffe s. str. PC1 separates Nubian and reticulated from Masai giraffe s. str., and PC2 separates Nubian from reticulated giraffe. The PCA space is further explored in Additional file 2: Fig. S2. (**c**) Ancestry proportions estimated from the same SNP dataset for $K = 3$ and $K = 9$. Colors indicate an individual's cluster membership. The numbers in between plots represent sampling locations according to (a). Interspecies admixture is found mostly between Nubian and reticulated giraffe at $K = 3$ and is restricted to two individuals from Loisaba Conservancy at $K = 9$. Admixture analyses for $K = 2$–11 are shown in Additional file 2: Fig. S3. (**d**) Mean likelihood and standard error (SE) across 100 runs per K. Mean likelihoods start plateauing at $K = 3$. (**e**) Assessment of admixture model fit based on the correlation of residuals for $K = 3$ and $K = 9$. Plotted values are the mean correlation within and between individuals from each sampling locality. Model fit assessments for $K = 1$–11 showing the pairwise correlation of residuals between all individuals are available in Additional file 2: Fig. S4. The order of sampling localities is the same as in (c). Localities with only one sampled individual are shown in grey.

Fig. 2. Nuclear and mitochondrial phylogenomic relationships among giraffe. Maximum likelihood phylogenies estimated from (**a**) 364,675 SNPs from 125 giraffe and (**b**) 13 mitochondrial protein-coding genes from 146 giraffe. The okapi was used as an outgroup (not shown). Colored tip labels indicate taxonomic assignment. Highly supported nodes (UFboot2 ≥ 95 and SH-aLTR ≥ 80) are marked with a black circle. In the nuclear tree, individuals formed clades corresponding to their respective species with high support. Mitochondrial introgression is observed from reticulated to Masai and from Masai to South

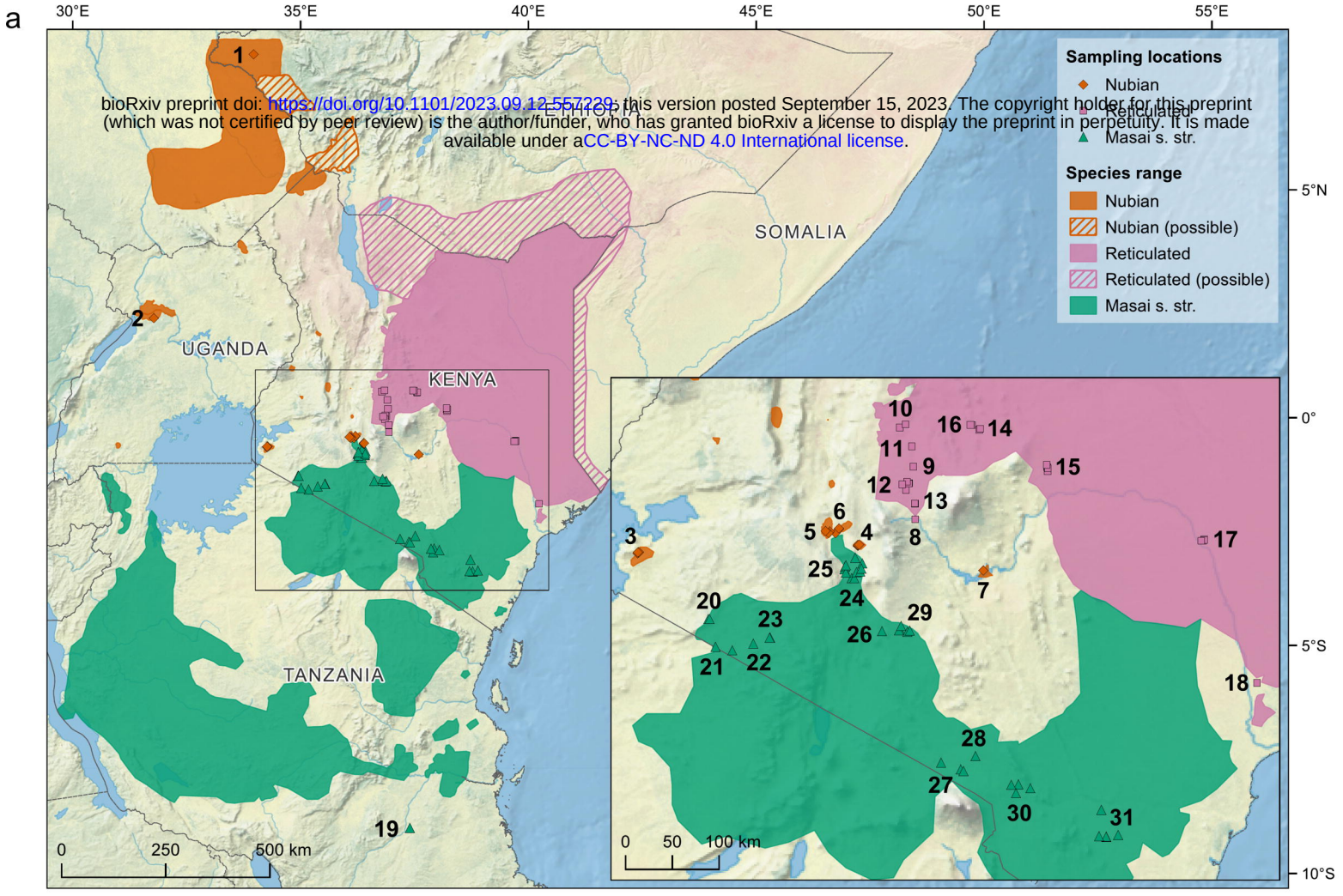
755 African giraffe. Individual GF292 carries a Nubian giraffe mitochondrion and falls between
756 the northern giraffe (i.e., West African, Kordofan, and Nubian) and the reticulated giraffe
757 clades in the nuclear phylogeny; thus, likely representing a natural hybrid.

758 **Fig. 3. Signatures of gene flow among Nubian, reticulated, and Masai giraffe s. str.**
759 **populations.** (**a**) Admixture graph of giraffe populations with migration events. The okapi
760 was used as an outgroup but was omitted from the image for a better resolution of the
761 divergence between giraffe populations. Nubian, reticulated, and Masai s. str. giraffe
762 populations were defined following the best fit admixture model (K = 9; Fig. 1c–e). Numbers
763 within parentheses in Nubian, reticulated, and Masai s. str. giraffe population labels indicate
764 sampling locations according to Fig. 1a. Migration arrows are colored according to their
765 weight and marked following the number of migration events (m = 1 or m = 2) allowed in the
766 model in which they were first inferred. The complete admixture graphs inferred for m = 1
767 and m = 2 and their corresponding residual fits are shown in Additional file 2: Fig. S6. (**b**)
768 Heatmap showing the *f*-branch (f_b) statistic estimated based on the topology recovered by
769 OrientAGraph. f_b values are shown for tests where the p-value of the associated D statistic is
770 < 0.01. Gray boxes indicate tips/branches which cannot be tested under a ((P1, P2) P3,
771 Outgroup) topology. (**c**) Contemporary migration rates among Nubian, reticulated, and Masai
772 giraffe s. str. Posterior mean migration rates were estimated based on 8,137 randomly
773 sampled unlinked SNPs from 97 wild giraffe. Links with arrow tips indicate migration
774 direction. Link widths are proportional to the fraction of individuals in the recipient population
775 with ancestry in the source population (per generation). Scale ticks represent the cumulative
776 fraction of migrants (per generation). Posterior estimates and 95% credible sets for migration
777 rates are provided in Additional file 3: Table S2.

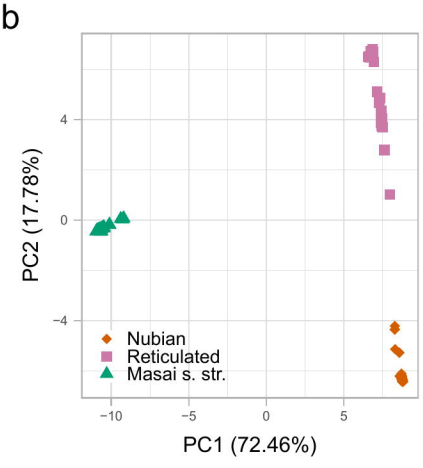
778 **Fig. 4. Demographic history of Nubian, reticulated, and Masai giraffe s. str.** Population
779 size changes over the recent past were reconstructed from the site frequency spectrum
780 (SFS) using the stairway plot model after masking singletons. Axes were scaled by a
781 mutation rate of 2.12×10^{-8} substitutions per site per generation and a generation time of 10

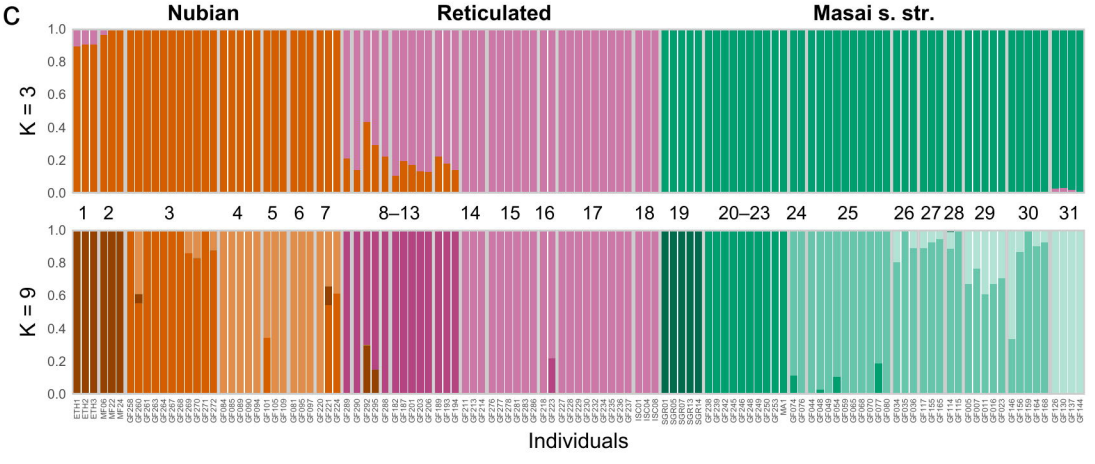
782 years. Colors represent focal giraffe taxa. Solid lines indicate median N_e estimates and

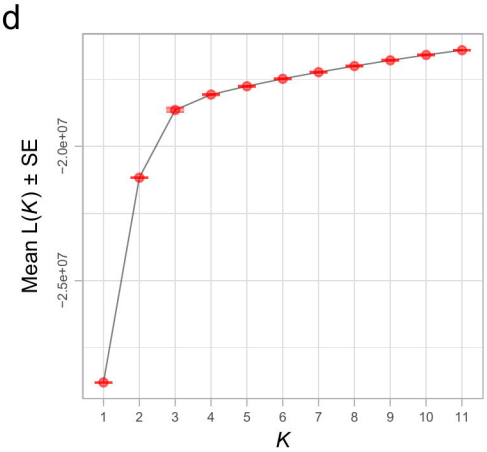
783 shaded areas correspond to 95% confidence intervals.

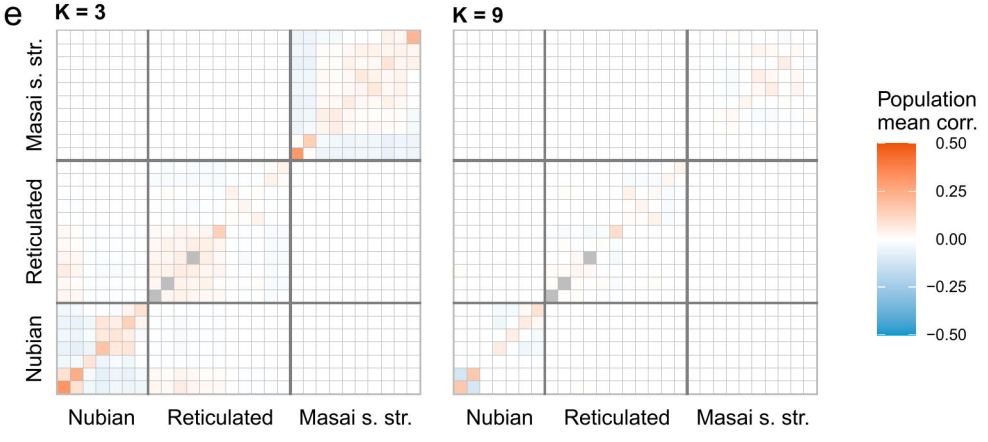


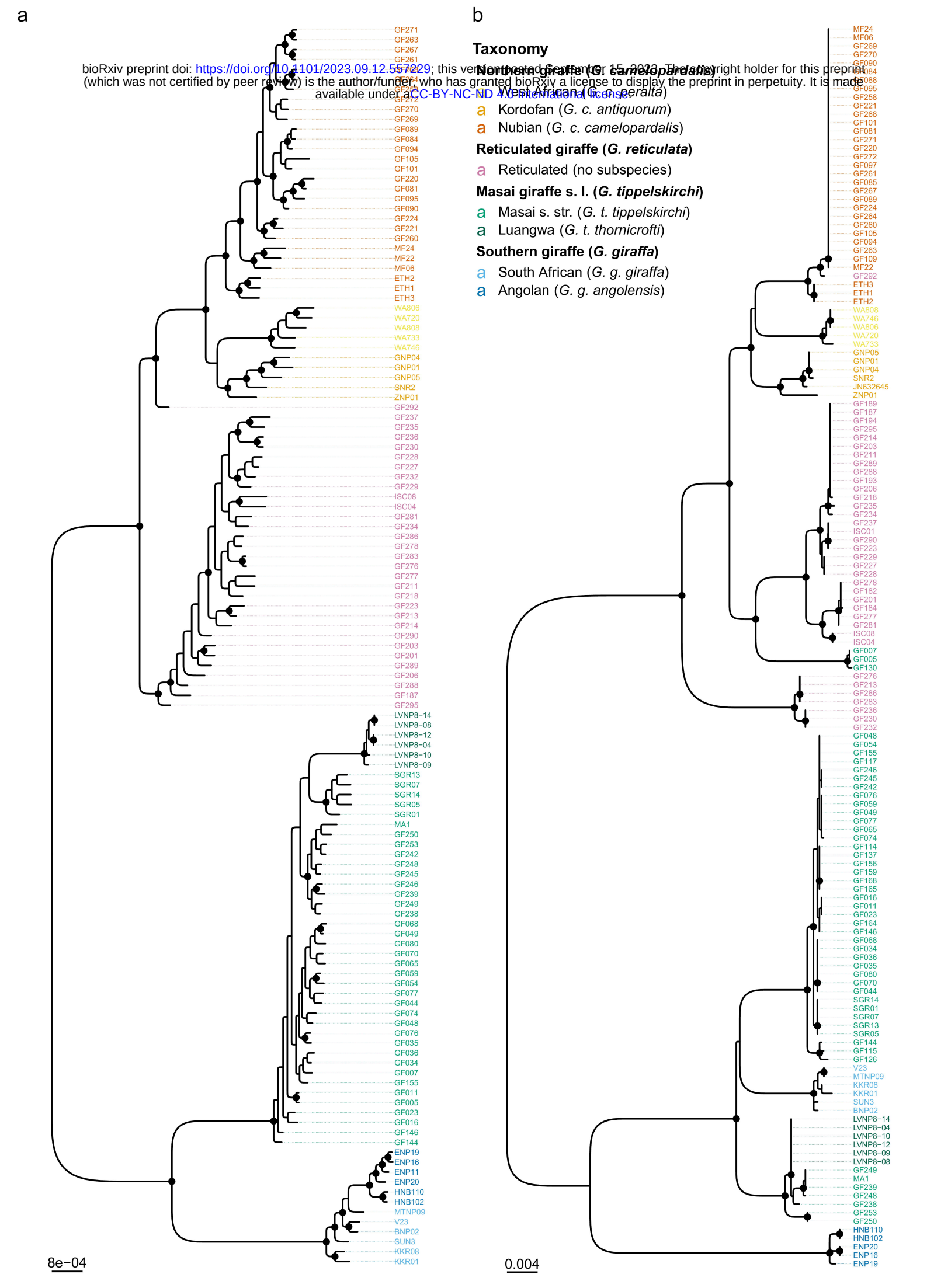
1. Gambella National Park
2. Murchison Falls National Park
3. Ruma National Park
4. Kigio Wildlife Conservancy
5. Lake Nakuru National Park
6. Soysambu Conservancy
7. Mwea National Reserve
8. Laikipia (Aberdare Country Club)
9. Laikipia (El Karama Conservancy)
10. Laikipia (Loisaba Conservancy)
11. Laikipia (Mpala Conservancy)
12. Laikipia (Ol Pejeta Conservancy)
13. Laikipia (Solio Game Reserve)
14. Buffalo Springs National Reserve
15. Meru National Park
16. Samburu National Reserve
17. Garissa
18. Ishaqbini Conservancy
19. Selous Game Reserve
20. Masai Mara (Kawai)
21. Masai Mara National Reserve
22. Masai Mara (Mara Bushtops Conservancy)
23. Masai Mara (Olarro Conservancy)
24. Hell's Gate National Park
25. Naivasha (private ranches)
26. Ngong
27. Amboseli National Park
28. Mbirikani
29. Nairobi National Park
30. Tsavo West National Park
31. Tsavo East National Park

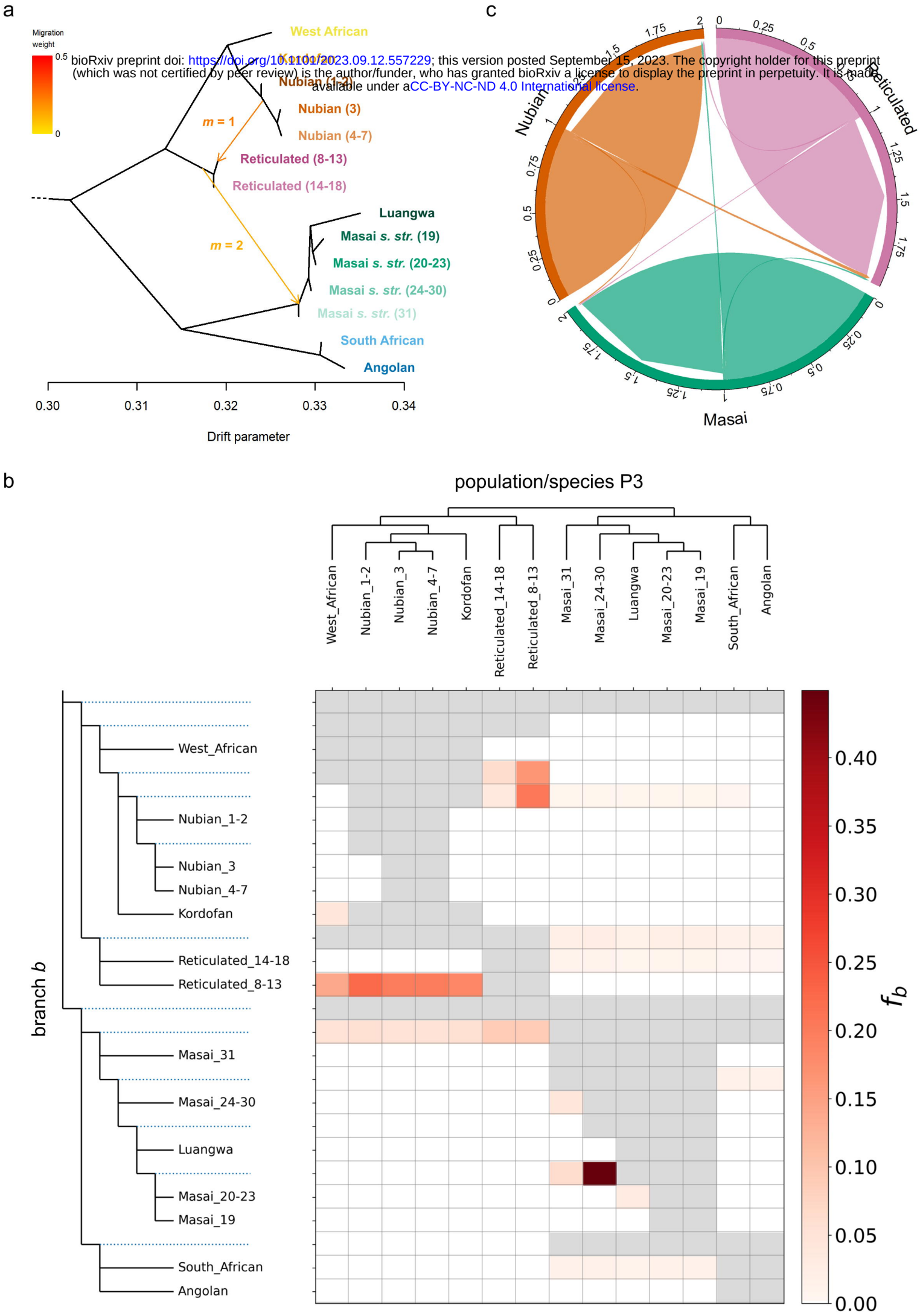


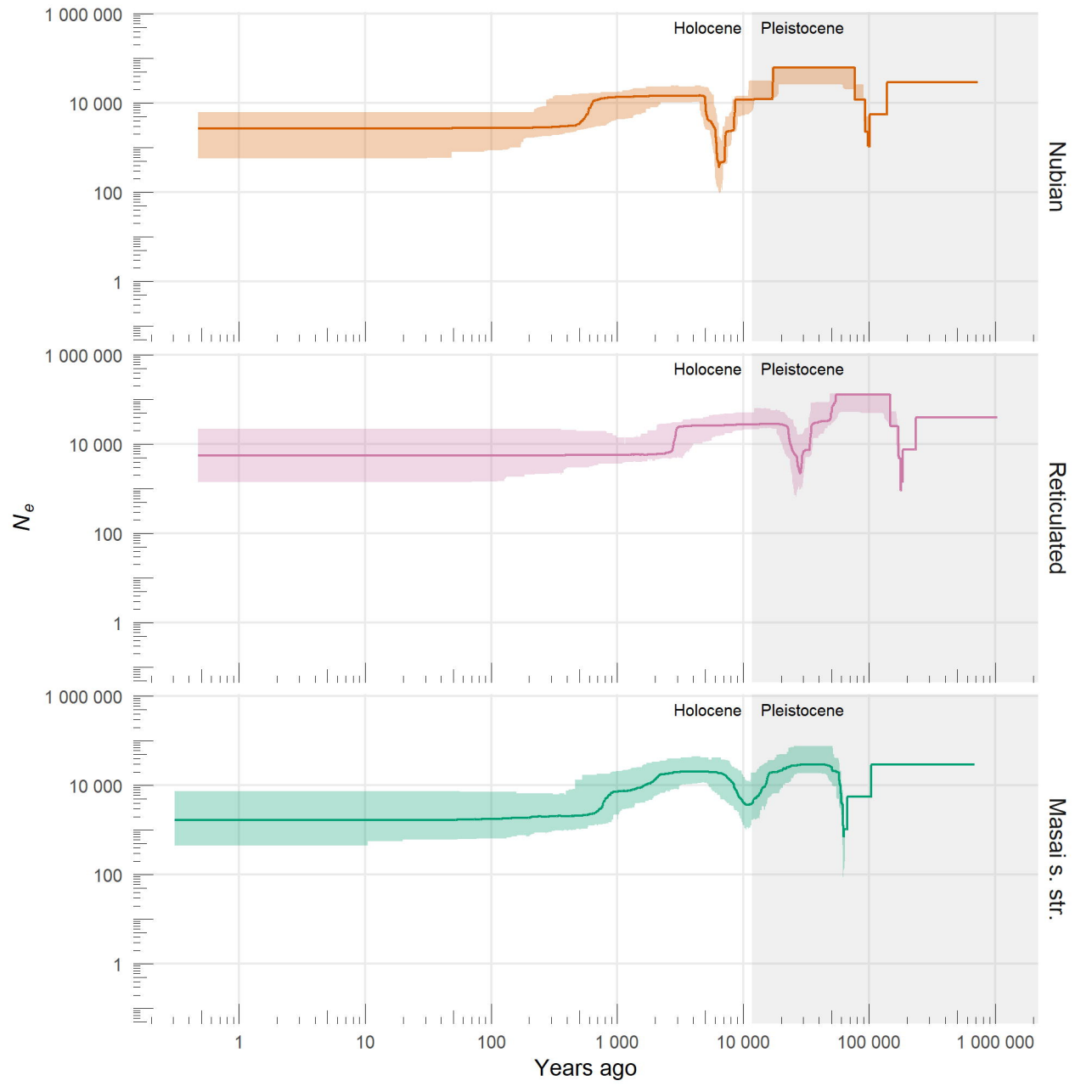











Holocene
Pleistocene
1 000 000
10 000
100
1
Nubian
Reticulated
Masai s. str.
N_e
1
10
100
1 000
10 000
100 000
1 000 000
Years ago