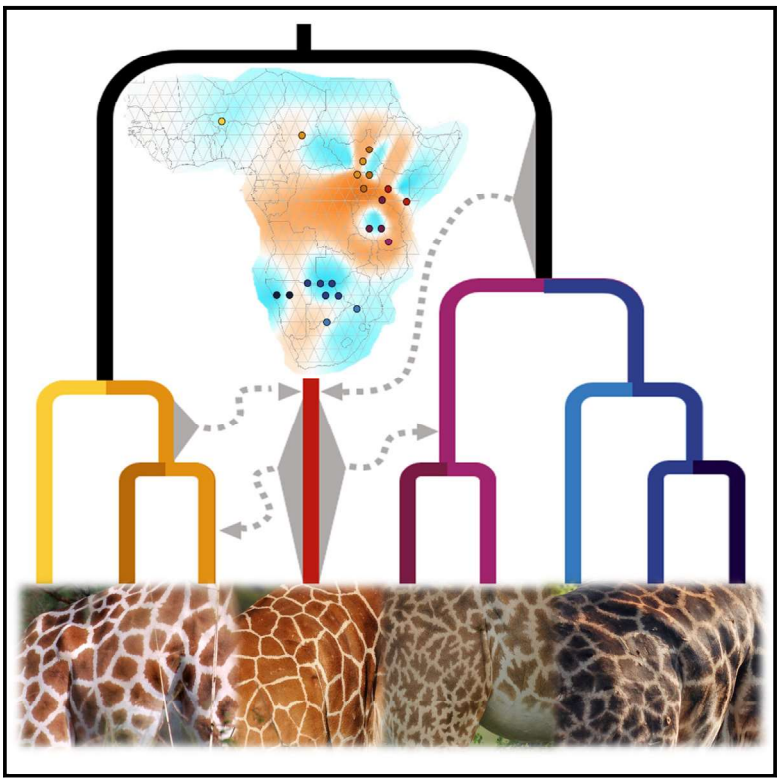
Report

Current Biology

Giraffe lineages are shaped by major ancient admixture events

Graphical abstract



Authors

Laura D. Bertola, Liam Quinn, Kristian Hanghøj, ..., Ida Moltke, Anders Albrechtsen, Rasmus Heller

Correspondence

ida@bio.ku.dk (I.M.), aalbrechtsen@bio.ku.dk (A.A.), rheller@bio.ku.dk (R.H.)

In brief

Giraffes consist of four major lineages, which show strong divergence despite being geographically close. Following analyses of whole genomes from 90 wild giraffes from throughout their range, Bertola et al. show that the evolutionary history of giraffes is complex and marked by major gene flow, in particular affecting Reticulated giraffes.

Highlights

- Giraffes show exceptional genetic structure despite lack of physical barriers
- Giraffes have a complex evolutionary history, with high levels of gene flow
- Reticulated giraffes are a hybrid lineage
- For effective conservation, diversity within giraffes needs to be taken into account

Bertola et al., 2024, Current Biology *34*, 1576–1586
April 8, 2024
https://doi.org/10.1016/j.cub.2024.02.051

CellPress

CellPress
OPEN ACCESS

Current Biology

Report

Giraffe lineages are shaped by major ancient admixture events

Laura D. Bertola,[1,9,11] Liam Quinn,[1,11] Kristian Hanghøj,[1,11] Genís Garcia-Erill,[1] Malthe Sebro Rasmussen,[1] Renzo F. Balboa,[1] Jonas Meisner,[1] Thomas Bøggild,[1] Xi Wang,[1] Long Lin,[1] Casia Nursyifa,[1] Xiaodong Liu,[1] Zilong Li,[1] Mumbi Chege,[2,3] Yoshan Moodley,[4] Anna Brüniche-Olsen,[1] Josiah Kuja,[1] Mikkel Schubert,[5] Morris Agaba,[6] Cindy G. Santander,[1] Mikkel-Holger S. Sinding,[1] Vincent Muwanika,[7] Charles Masembe,[8] Hans R. Siegismund,[1] Ida Moltke,[1,*] Anders Albrechtsen,[1,*] and Rasmus Heller[1,10,12,*]

[1]Department of Biology, University of Copenhagen, Copenhagen, Denmark
[2]Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands
[3]Wildlife Research and Training Institute, Naivasha, Kenya
[4]Department of Biological Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa
[5]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark
[6]School of Life Sciences and Bioengineering, Nelson Mandela African Institution of Science and Technology, Nelson Mandela Road, Arusha, Tanzania
[7]College of Agricultural and Environmental Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda
[8]College of Natural Sciences, Makerere University, P O. Box 7062, Kampala, Uganda
[9]X (formerly Twitter): @LauraDBertola
[10]X (formerly Twitter): @popgenDK
[11]These authors contributed equally
[12]Lead contact
*Correspondence: ida@bio.ku.dk (I.M.), aalbrechtsen@bio.ku.dk (A.A.), rheller@bio.ku.dk (R.H.)
https://doi.org/10.1016/j.cub.2024.02.051

SUMMARY

Strong genetic structure has prompted discussion regarding giraffe taxonomy,[1–3] including a suggestion to split the giraffe into four species: Northern (*Giraffa c. camelopardalis*), Reticulated (*G. c. reticulata*), Masai (*G. c. tippelskirchi*), and Southern giraffes (*G. c. giraffa*).[4–6] However, their evolutionary history is not yet fully resolved, as previous studies used a simple bifurcating model and did not explore the presence or extent of gene flow between lineages. We therefore inferred a model that incorporates various evolutionary processes to assess the drivers of contemporary giraffe diversity. We analyzed whole-genome sequencing data from 90 wild giraffes from 29 localities across their current distribution. The most basal divergence was dated to 280 kya. Genetic differentiation, F_{ST}, among major lineages ranged between 0.28 and 0.62, and we found significant levels of ancient gene flow between them. In particular, several analyses suggested that the Reticulated lineage evolved through admixture, with almost equal contribution from the Northern lineage and an ancestral lineage related to Masai and Southern giraffes. These new results highlight a scenario of strong differentiation despite gene flow, providing further context for the interpretation of giraffe diversity and the process of speciation in general. They also illustrate that conservation measures need to target various lineages and sublineages and that separate management strategies are needed to conserve giraffe diversity effectively. Given local extinctions and recent dramatic declines in many giraffe populations, this improved understanding of giraffe evolutionary history is relevant for conservation interventions, including reintroductions and reinforcements of existing populations.

RESULTS

Motivation and aims

In many species and species groups there is no obvious bifurcating tree that describes taxonomic relationships satisfactorily,[7] and previous studies have shown that interspecific introgression is common and may even play a role in speciation in multiple mammalian lineages,[8] including canids,[9] felids,[10,11] and bears.[12] Giraffes (*Giraffa* sp.) provide an interesting case study, as these widely dispersed megaherbivores have been described as strongly diverged yet with a contested taxonomy.[1] Although the IUCN (International Union for Conservation of Nature) only recognizes a single species with nine subspecies,[13] in recent years there have been proposals to classify giraffes into four separate species.[4–6] High levels of genetic differentiation, even between directly neighboring populations, have been central in this newly suggested classification. However, others have pointed out that different genetic markers do not show congruent results and that, depending on the methodology used, one, three, or six species could be distinguished.[1–3]

Historically, giraffes had a continent-wide range throughout savannah habitats in sub-Saharan Africa; however, their range

has recently become severely fragmented as a result of anthropogenic habitat conversion and degradation.[14,15] Giraffe populations are estimated to have declined 40% in the past 30 years, restricting them mostly to conservation areas.[15,16] Assessing both historical and contemporary gene flow will help understand unresolved phylogeographic patterns but can also be used to inform conservation actions. This is particularly relevant since giraffes are frequently translocated,[17] and insights into genetic relationships should be taken into account when selecting source and target populations for conservation translocations.[18,19] Motivated by this, we aimed to explore the evolutionary and demographic history of giraffes using whole-genome data from across the range, covering all previously described major lineages.

Sequencing and quality control

We generated medium depth (≈20X) whole-genome sequencing data for 47 giraffes from 12 localities in six countries (Table S1). In addition, we obtained published whole-genome sequencing data for 43 other giraffes from 17 localities in 11 countries,[4,20] as well as a giraffe[21] and an okapi (*Okapia johnstoni*)[20,22] reference genome and sequencing data from another okapi[20,22] and a pronghorn (*Antilocapra americana*)[22] as outgroups. After mapping, 11 samples were excluded due to low average depth (<6X) or highly skewed, i.e., non-symmetric, depth distributions. One individual from a pair of close (first degree) relatives was also removed, leaving 78 samples spanning 27 localities in 13 countries and all four major lineages (Figure 1A). Excluded samples are listed in Table S1 (light gray shading).

Population structure, differentiation, and connectivity

To explore the evolutionary history of giraffes, we first inferred population structure using a combination of methods, including ADMIXTURE[23] and Principal Component Analysis (PCA).[24] Estimated admixture proportions are shown in Figure 1B for both K = 4 (upper barplot) and for K = 10 (bottom barplot) (see Figure S1 for K = 2 to K = 10 and the correlations of residuals from EvalAdmix[25]). These two levels of structure highlight the hierarchical distribution of genetic diversity in giraffes. K = 4 gives rise to a clean separation into the four major lineages that correspond with proposed species according to some recent studies[4–6] (but see Bercovitch et al.[1,2] and Petzold and Hassanin[3]): Northern (*Giraffa c. camelopardalis*), Reticulated (*G. c. reticulata*), Masai (*G. c. tippelskirchi*), and Southern (*G. c. giraffa*) giraffe. However, the pairwise correlation of residuals from EvalAdmix[25] suggested the presence of underlying population structure within these major lineages, which is largely resolved when using K = 10, although EvalAdmix suggested that there may still be unresolved structure in the Nubian and Angolan sublineages, driven by a group of possible second-degree relatives or unresolved population sub-structure. Furthermore, sublineage structure appeared in the Reticulated lineage as a result of four related individuals that formed their own cluster (indicated with gray asterisks in Figure 1B). When we removed these four individuals, there were nine groups remaining, roughly corresponding to previously described subspecies (see Figure S1 for K = 9 excluding these relatives): the Northern giraffe was split into three sublineages, West African (*G. c. peralta*), Kordofan (*G. c. antiquorum*), and Nubian (*G. c. camelopardalis*); the Masai giraffe (*G. c. tippelskirchi*) was split into two sublineages, Masai and Masai-Selous; and the Southern giraffe was split into three sublineages, Southern African (*G. c. giraffa*), Southern Central (*G. c. giraffa*), and Angolan (*G. c. angolensis*) (for corresponding sampling locations see Figure 1A). These groupings were confirmed by PCA (Figure S2), in which the first two components separate the four major lineages, with Reticulated clustering close to the Northern giraffe. Substructure within the four lineages becomes apparent in the following principal components, with Reticulated giraffes splitting out on PC3. This is consistent with the pattern described in Coimbra et al..[4] A neighbor-joining tree based on pairwise Identity-by-State (IBS) values also clearly illustrated the substructure within all four major lineages except for the Reticulated giraffe (Figure 1C).

Based on these results, we consider all Reticulated giraffes in our dataset as members of a single sublineage for downstream analyses, leaving a total of nine biologically meaningful sublineages for exploring giraffe evolutionary history. Since many of the downstream analyses require units to be homogeneous, we used these nine sublineages to identify individuals with signs of significant recent admixture. Five individuals with >10% admixture proportions (indicated with black asterisks [*] in Figure 1B) were excluded from all downstream analyses for which discrete populations were required. We then checked for non-homogeneity within each of the nine sublineages driven by outlier individuals using D-statistics (ABBA-BABA).[26] For each sublineage we tested all pairs of individuals from the relevant sublineage (H1 and H2), an individual from another sublineage within the same major lineage (using a Nubian individual for Reticulated giraffes) (H3), and okapi as an outgroup (H4). Two individuals, GCamTzW_3369 (Masai) and GCamZwS_1546 (Southern African), consistently showed deviations from homogeneity (Figures S3A and S3B) and were subsequently removed from further analyses. Thereafter, the distribution of test scores was centered on zero, and the majority of tests did not deviate significantly (Figure S3B), confirming the overall genetic homogeneity of the assumed sublineages in our study. Finally, we derived linkage disequilibrium (LD) decay curves for each of the nine sublineages (Data S1A), and for the Angolan and Masai-Selous giraffes we observed that these curves plateau at a higher level than for the other populations. This could be a sign of recent admixture or substructure, which can confound our ability to infer ancient gene flow and other historical events in downstream analyses.

Genetic differentiation between the four major lineages, calculated as pairwise Hudson F_{ST},[27] ranges from 0.28 (Northern-Reticulated) to 0.62 (Northern-Southern) (Figure 2A, upper matrix). Pairwise F_{ST} between each of the nine sublineages (Figure 2A, bottom matrix) ranges between 0.10 (Masai - Masai-Selous; Southern African-Southern Central) and 0.71 (West African-Angolan; Kordofan-Angolan), with 0.24 being the highest value within a major lineage (West African-Nubian). To place the population structure in a geographical context, we performed an analysis of the spatial distribution of genetic distances using Estimating Effective Migration Surfaces (EEMS).[28] We observed apparent barriers to gene flow throughout the Central African rainforest, which is an unsuitable habitat for giraffes, as well as along the geographic boundaries of the four major lineages (Figures 2B and 2C), i.e., between Reticulated, Nubian, and Masai sublineages. In addition, less pronounced barriers were

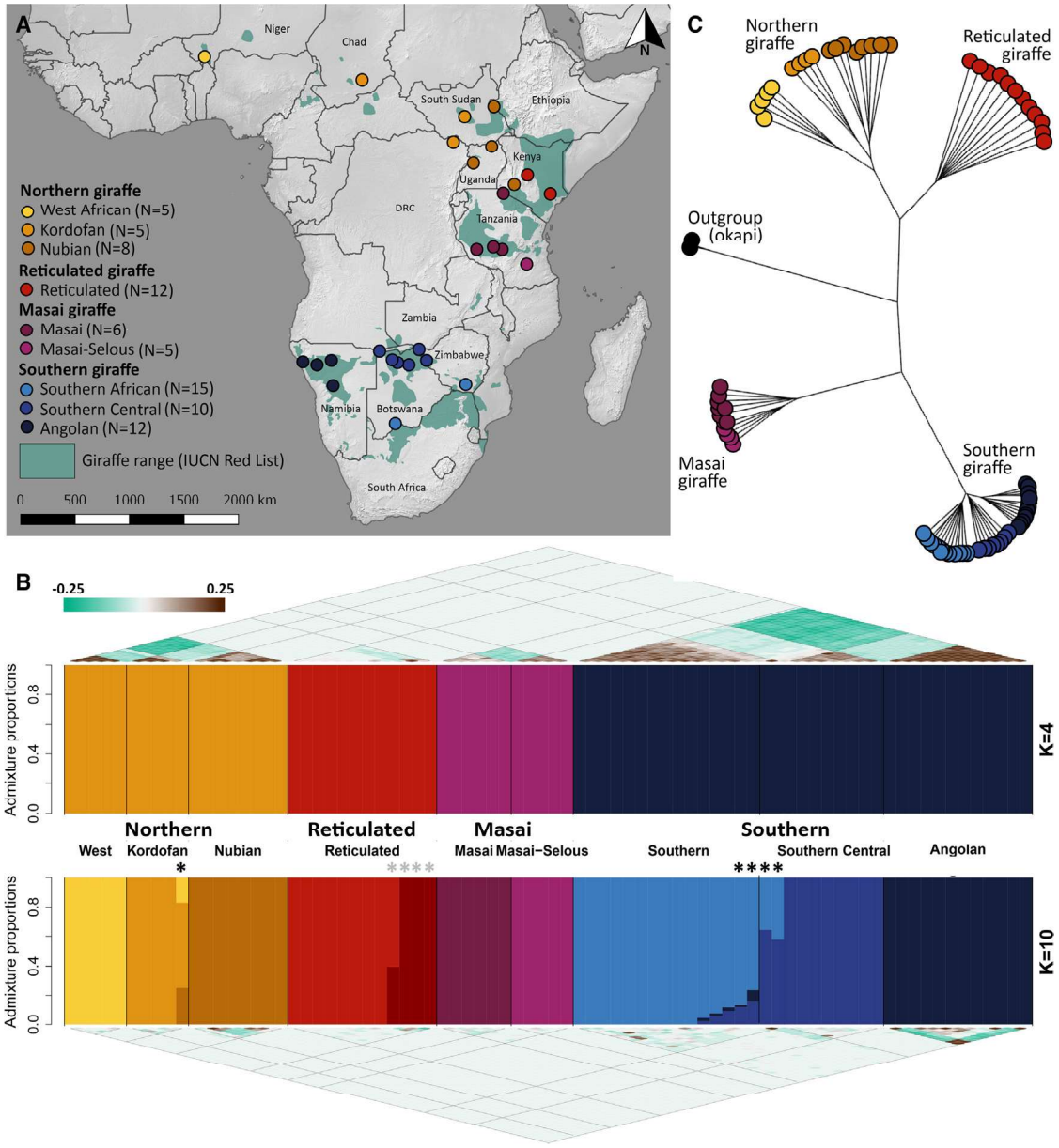


Figure 1. Sampling and population structure of giraffes

(A) Distribution of giraffe samples used in this study, colored according to inferred sublineages. Contemporary giraffe distribution, fragmented from a formerly more continuous distribution is shown in green.[13]

(B) ADMIXTURE results assuming K = 4 (upper barplot) and K = 10 (bottom barplot). EvalAdmix results, displaying the correlation of residuals for assessing model fit from ADMIXTURE, are shown as correlation plots above the K = 4 and below the K = 10 barplot. Gray asterisks (*) highlight the related individuals within the Reticulated giraffe, and black asterisks indicate samples that have been excluded due to admixture signal (>10% ancestry proportion). Also see Figure S1.

(C) Neighbor-joining tree based on pairwise Identity-by-State (IBS), illustrating the substructure within the major lineages.

inferred among the Nubian, Kordofan, and West African sublineages, within the Northern giraffe, and between Southern giraffe sublineages (Figures 2B and 2C). In most cases these barriers correlate with geographical features, such as the Gregory Rift Valley (Northern-Masai split) and the Kalahari Desert (Angolan-Southern Central split), and several permanently flowing river systems including the Shari-Logone (West-Kordofan), Nile (Nubian-Masai/Reticulated; but note that Nubian giraffes are also found in Kenya, east of the Nile), Tana (Masai-Reticulated), Zambezi (Masai-Southern), and Kavango (Angolan-Southern Central).

Heterozygosity and demographic history

Per individual genome-wide heterozygosity estimates showed that the Reticulated lineage had the highest genome-wide heterozygosity values ($H = 0.00197$). The three Northern giraffe sublineages also showed relatively high heterozygosity ($H = 0.00137–0.00143$) with a gradual decline toward the Southern

A	Northern			Reticulated	Masai		Southern		
Northern	-			-	-		-		
Reticulated	0.28			-	-		-		
Masai	0.55			0.51	-				
Southern	0.62			0.58	0.54		-		
West African	-	-	-	-	-	-	-	-	-
Kordofan	0.24	-	-	-	-	-	-	-	-
Nubian	0.24	0.16	-	-	-	-	-	-	-
Reticulated	0.35	0.33	0.32	-	-	-	-	-	-
Masai	0.64	0.63	0.62	0.55	-	-	-	-	-
Masai-Selous	0.67	0.66	0.65	0.58	0.10	-	-	-	-
Southern African	0.70	0.69	0.68	0.62	0.63	0.67	-	-	-
Southern Central	0.69	0.68	0.67	0.60	0.62	0.66	0.10	-	-
Angolan	0.71	0.71	0.70	0.63	0.66	0.69	0.19	0.14	-
	West African	Kordofan	Nubian	Reticulated	Masai	Masai-Selous	Southern African	Southern Central	Angolan

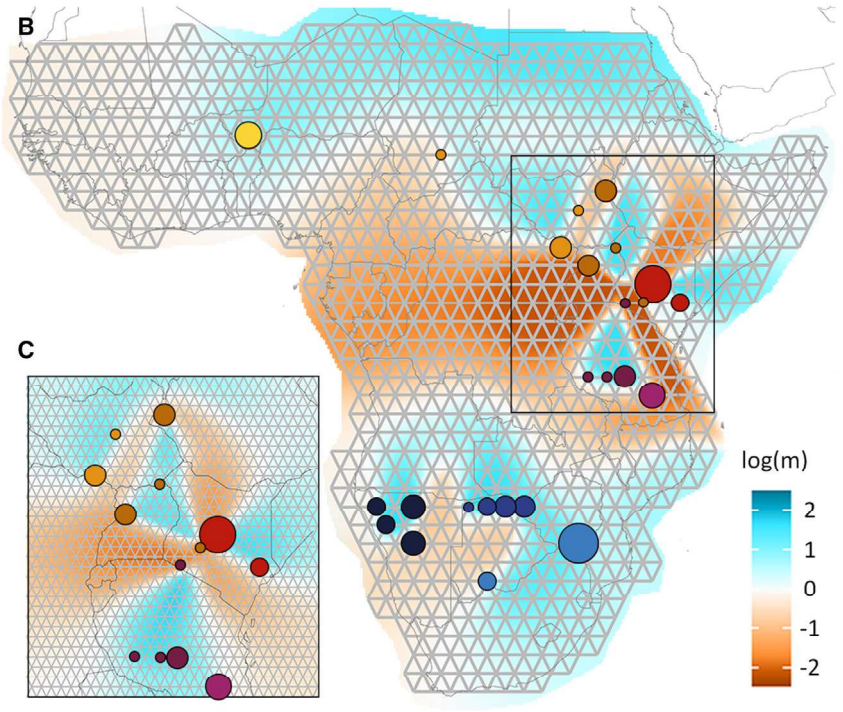


Figure 2. Genetic differentiation and migration rates of giraffes

(A) Table showing genetic differentiation between sublineages, expressed by pairwise Hudson F_{ST} derived from the site frequency spectrum (SFS).

(B) EEMS plot showing migration rates (log(*m*)) between sampling localities across the giraffe range, with high log(*m*) rates indicating increased gene flow and low log(*m*) rates indicating reduced gene flow under the null model of isolation by distance.

(C) EEMS plot showing migration rates inferred at a local scale in eastern Africa (black box), where three major giraffe lineages occur but appear geographically isolated.

sublineages (H = 0.00060–0.00077) (Figure 3A, showing both genome-wide heterozygosity and heterozygosity adjusted for runs of homozygosity [ROH], see STAR Methods). We note that our estimated values are about three times higher than previously reported estimates.[4] We found that this difference is likely due to the use of BCFTools v1.9,[29] where the original base alignment quality (BAQ) algorithm leads to an underestimation of heterozygous positions as the algorithm was designed for datasets that do not involve potentially problematic alignments (Data S1B). The heterozygosity pattern in which Reticulated giraffes show the highest diversity and in which there is a gradual decline towards the southern sublineages is mirrored by a gradual increase in the fraction and length of ROH (Figure 3B). Although there are pronounced differences between individuals within the same sublineage, it appears that ROH occur in higher frequency in the Southern lineage, in line with results previously reported by Coimbra et al.[4] According to Coimbra et al.,[4] Masai giraffes also display a relatively high proportion of the genome in ROH; however, by including more samples from other Masai populations, we show that there can be substantial differences between populations within the same lineage. ROH occur less frequently and in shorter lengths in the Reticulated lineage, also reported by Coimbra et al.,[4] possibly as a result of introgression from other major lineages (see below).

Historical population sizes inferred by the Pairwise Sequentially Markovian Coalescent (PSMC)[30] show overall higher historical population sizes for sublineages in the Northern giraffe, compared to the Masai and Southern giraffe (Figure 3C, and see Data S1C for a comparison of PSMC results with and without masking of ROH, which shows that the PSMC results are not affected by the ROH except for in the most recent time period, where PSMC is known to be unreliable[31]). This is in line with results previously reported in Coimbra et al.[4] The relatively sharp increase in Reticulated giraffe around 110 kya, also detected in Coimbra et al.,[4] may not reflect changes in actual population size, but rather the result of introgression from other lineages (see below), leading to loci with longer coalescent times, elevated heterozygosity, and decreased ROH. Split times estimated using the Two-Two (TT) method[32] date the oldest divergence in our dataset, i.e., the most recent common ancestor between the four major lineages, at around 280 kya (Data S1D), which is also in line with the separation of populations in PSMC. This estimate also falls within the range of split times Coimbra et al.[4] derived from a protein-coding dataset, where they report 230–370 kya. Fastsimcoal[33] (see below) infers a slightly younger split, at ≈200 kya. TT and fastsimcoal agreed that the split time between the Masai and Southern lineages is older (TT: 70 kya, fastsimcoal: 106 kya) than the split between the Northern sublineages (TT: negative values,

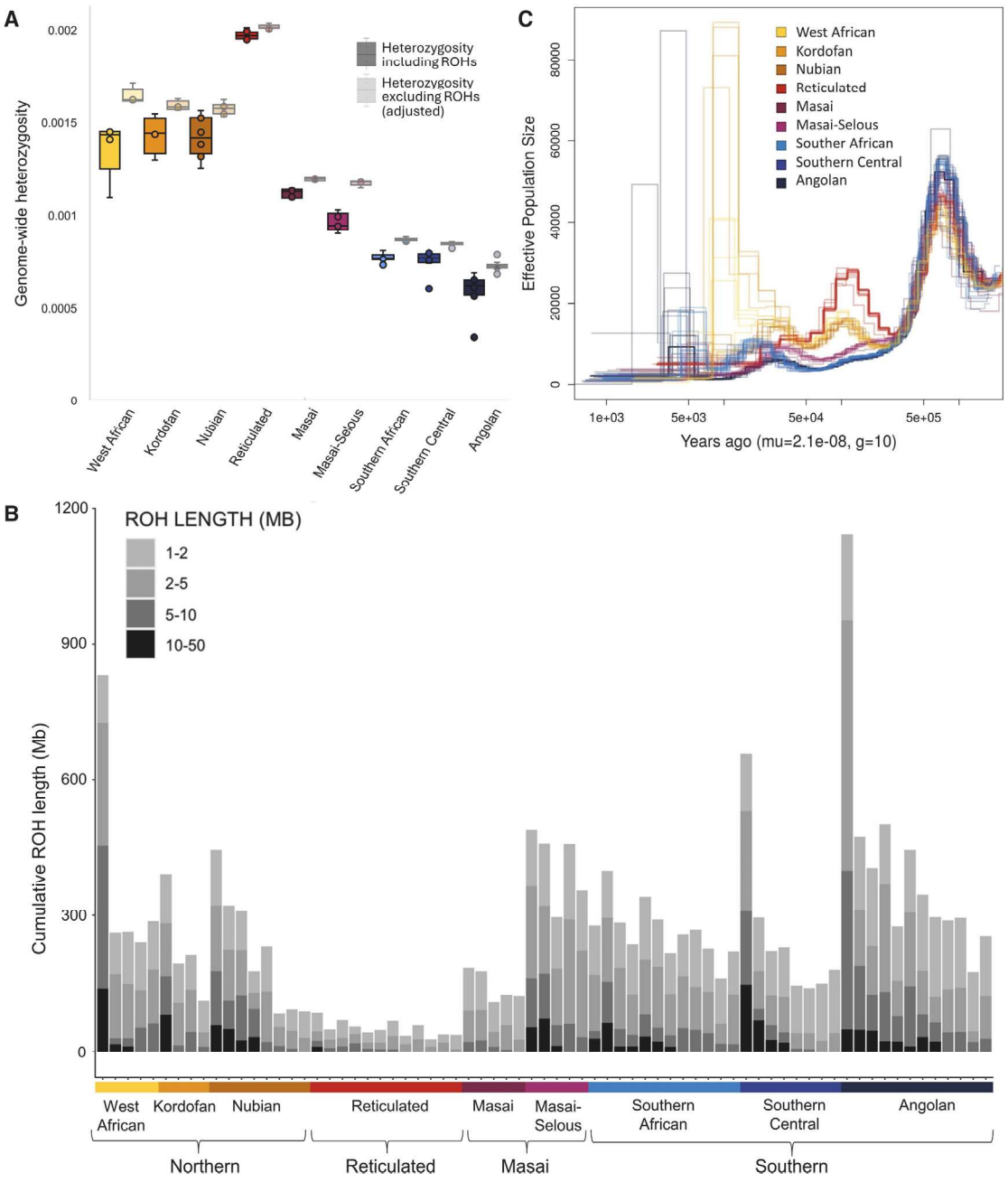


Figure 3. Heterozygosity, cumulative length of runs of homozygosity (ROH), and effective population size reconstructed by Pairwise Sequentially Markovian Coalescent (PSMC)

(A) Heterozygosity estimates for 71 giraffes clustered into nine sublineages, excluding admixed individuals (>10% admixture proportions). Boxplots with saturated colors reflect heterozygosity estimates including ROH, while grayed out boxes refer to "adjusted heterozygosity," i.e., corrected for the presence of ROH. Analyses were executed on dataset 2 (all genotypes).

(B) Cumulative length of ROH in 71 giraffes from nine sublineages, shown for ROH in various length bins (gray shading), ranging from 1 to 50 MB. We excluded admixed individuals (>10% admixture proportions) and two individuals identified by D-statistics as deviating from a homogeneous sublineage (Figure S3). Analyses were executed on dataset 1 (all SNPs).

(C) Effective population size reconstructed by PSMC of 71 giraffe samples, representing nine sublineages. We excluded admixed individuals (>10% admixture proportions) and two individuals identified by D-statistics as deviating from a homogeneous sublineage. For scaling, we used a mutation rate of $2.1 \cdot 10^{-8}$ per generation and a generation time of 10 years. Analyses were executed on dataset 2 (all genotypes). See also dataset 1.

fastsimcoal: 33–37 kya), although they differed in their estimates (Data S1D, Figures S4A and S4B), which is likely due to different modeling assumptions between the methods. The very recent and overlapping split times estimated by TT between each of the Northern sublineages and Reticulated giraffe could hint at gene flow between these two major lineages, whereas the reduced split times between the two Masai sublineages and Reticulated giraffes could also suggest that they were involved in gene flow.

Introgression and f-branch statistics

To investigate these potential signatures of introgression, f-branch statistics were calculated using Dsuite,[34] using a population topology inferred from the IBS-based neighbor-joining tree (Figure 1C). The topology of the tree is consistent with that inferred in a previous whole-genome analysis[4] and was further corroborated by the relative split times using the TT method[32] (Data S1D). F-branch statistics reflect correlations between allele frequencies among sets of populations, thereby giving information about shared ancestry and introgression. The f-branch results of our dataset show marked violations of the tree, indicated by the orange-red shading in the heatmaps (Figures 4A and 4C), which can only be explained by gene flow. Here, the rows (y axis) represent the sources and the columns (x axis) represent the recipients of introgression, where a darker shading indicates a higher level of gene flow. Note that ancestral lineages can also be a source, indicated by the dotted lines in the tree. For the giraffes, we find that in particular the Reticulated giraffe is flagged as being involved in ancestral gene flow events, with the two Masai sublineages and the Kordofan giraffe showing signs of admixture with the Reticulated lineage (Figures 4A and 4C, Data S1E). Further signs of gene flow were observed, e.g., within the Northern major lineage and between the two Masai sublineages and some of the Southern sublineages (Figure 4C). Having established that tree-likeness is violated at several different levels in the evolutionary history of giraffes, we next focused on inferring gene flow events between the major lineages. We first excluded the Angolan and Nubian sublineages to simplify the subsequent admixture graph modeling, as doing so removes some of the deviations from tree-likeness that were either subtle or indicative of within-lineage gene flow. We used findGraphs and QPgraph from AdmixTools2[35] to explore the space of possible admixture graphs involving zero to five admixture events. The results indicated that adding two admixture events significantly ($p < 0.05$) improved the fit compared to a model with one or no admixture events, while there was no further statistical improvement from admixture graphs with more than two admixture events (Figure 4B and Data S1F). For scenarios with two admixture events, we obtained two equally likely graphs, both suggesting that the Reticulated giraffe evolved through similarly high genetic contributions from a branch within the Northern major lineage and a basal branch of the Masai and Southern major lineages (60% and 40%, respectively). Considering the test scores (Data S1F), we conclude that an absolute minimum of two major admixture events involving all major lineages is necessary to account for cross-lineage allele sharing in a model that includes seven giraffe sublineages. We also explored admixture graphs based on all nine inferred sublineages (Figure 4D). Here, we found that even five admixture events are insufficient to explain the f-statistics among populations, and we conclude that more admixture events would be needed to fully explain the observed patterns at this level of complexity. Importantly, they also include the interlineage admixture involving the Reticulated giraffe suggested by the admixture graphs inferred without Angolan and Nubian giraffe sublineages. In addition, the best-supported graphs included several admixture events below the sublineage level and included minor (<5%) admixture events. When all nine sublineages were included in the graph (Figure 4D) a third cross-lineage admixture event was suggested, and numerous additional but unstable admixture events were suggested, with many equally likely admixture graphs. The graphs for two admixture events for the seven sublineages and the top graphs for three and five admixture events for the nine sublineages are shown in Data S1G. Results for the same workflow with giraffe data mapped to a giraffe reference genome are shown in Datas S1H and S1I.

To further corroborate admixture events in the context of population size changes and split times, we ran two different demographic models using fastsimcoal2 (Figures S4A and S4B). These results show that the Reticulated lineage has been strongly affected by admixture (54%) from the Kordofan sublineage, further confirming the hybrid character of Reticulated giraffes. In addition, it is also consistent with lower levels (6%) of subsequent introgression from Reticulated into Masai. The fastsimcoal estimates of the initial divergence time ranged between 198 and 212 kya (depending on model), slightly younger than the results derived from PSMC and TT analyses.

DISCUSSION

In this study, we present a comprehensive model of giraffe evolutionary history that specifically accounts for admixture events between major lineages and the populations within them. Most strikingly, we found that the Reticulated giraffe is essentially a hybrid lineage, formed by substantial genetic contributions from two currently highly divergent giraffe lineages. The exact nature and timing of this event, however, is difficult to determine as several admixture graphs fit the data equally well. In particular, we were neither able to determine the position on a branch when the gene flow event occurred (gray triangles, Figures 4B and 4D) nor explicitly test whether there was a ghost lineage involved. Demographic modeling suggested that the population donating ≈40% of the ancestry of the Reticulated giraffe could be a deeply divergent population branching off the southern main lineage close to the root of the giraffe tree, but we warn that this conclusion is tentative because it is contingent on the assumed model of, e.g., population size changes. Additionally, we could not distinguish whether the gene flow occurred in a single pulse or through multiple events spanning a longer period of time. The genetic structure we observed is largely consistent with previous studies[4,5,36]; however, we significantly improve on these through our detailed analyses of the role of ancient gene flow in giraffe evolution.[4,5] Our results are consistent with the emerging consensus that many evolutionary lineages have no obvious bifurcating tree that describes their taxonomic relationships satisfactorily.[7] Instead, patterns of genome-wide diversity can better be represented as complex networks that are often difficult to interpret.[37]

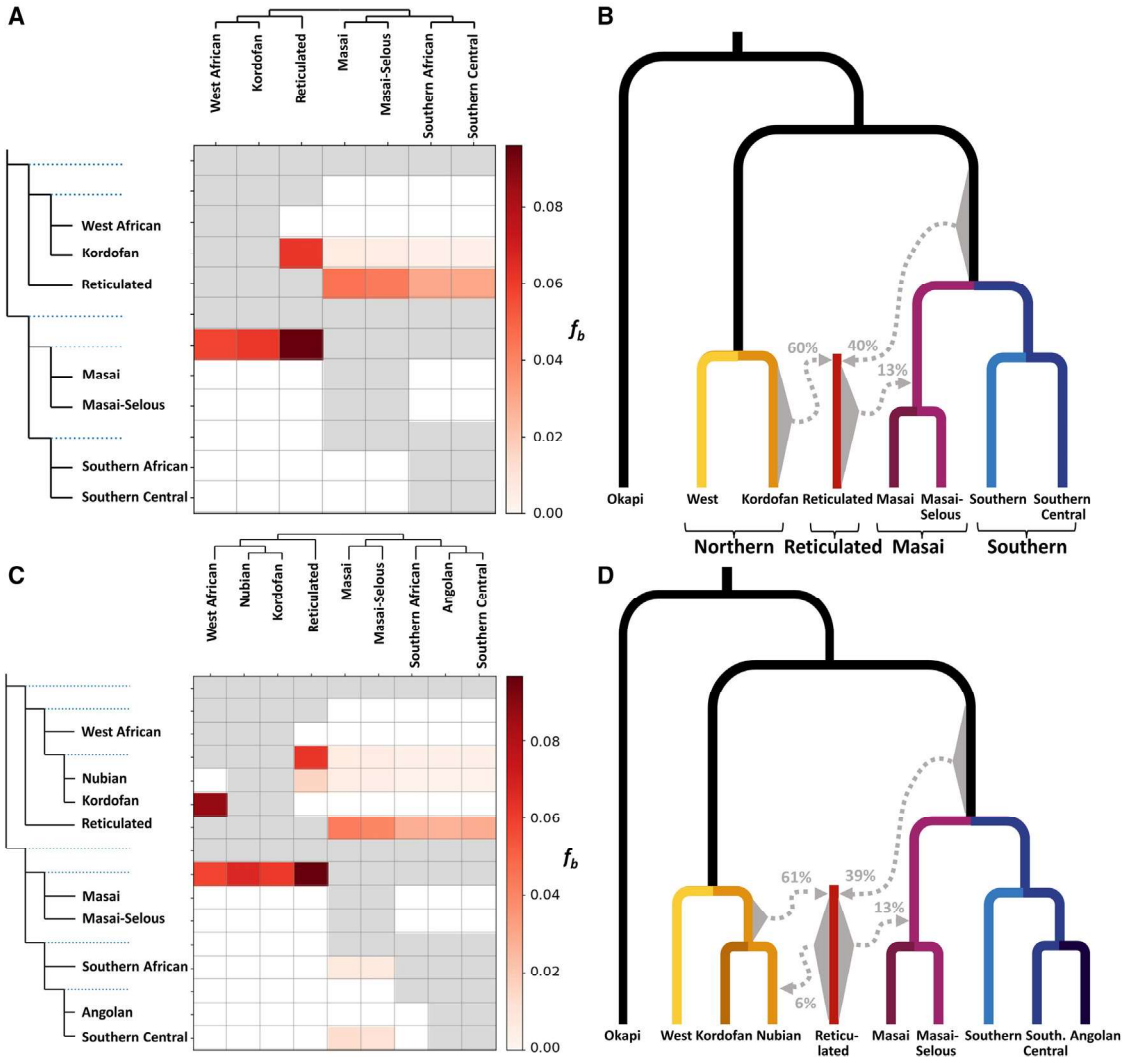


Figure 4. Complex evolutionary history of giraffes, including admixture between major evolutionary lineages based on data mapped to an okapi reference genome

(A) f-branch (f_b) heatmap for seven giraffe sublineages, excluding Nubian and Angolan, to reduce complexity as a result of intraspecific gene flow. The shading of the heatmap refers to excess allele sharing between branches located on the tree on the y axis (relative to its sister branch) and the sublineages identified on the x axis. Gray cells are combinations of branches and populations for which the f-branch cannot be calculated given the topology of the tree.

(B) Highest scoring admixture graph with two admixture events, including seven giraffe sublineages, exploring 0 to 5 admixture events. The gray triangles indicate that it is uncertain where along the branch these admixture events took place, as well as that introgression may not have occurred as a single pulse.

(C) f-branch (f_b) heatmap for all nine giraffe sublineages. The shading of the heatmap refers to excess allele sharing between branches located on the tree on the y axis (relative to its sister branch) and the sublineage identified on the x axis. Gray cells are combinations of branches and populations for which the f-branch cannot be calculated given the topology of the tree.

(D) Highest scoring admixture graph with three admixture events, including nine giraffe sublineages. Another version of this figure based on giraffe data mapped to a giraffe reference genome is included as Data S1H. Original admixture graphs, including drift parameters, for two, three, and five admixture events are available as Data S1G for data mapped to okapi and Data S1I for data mapped to giraffe.

The finding of extensive ancient admixture questions if and how reproductive barriers between major giraffe lineages have arisen, which is a requirement under some species concepts, notably the Biological Species Concept.[38] The admixture events we infer here occurred relatively early in giraffe evolutionary history, and it can be argued that reproductive isolation could have evolved more recently. However, we point out that giraffes from different major lineages readily hybridize in captivity[39] and possible hybrids have been documented from the wild,[40] although there is no indication for the existence of *in situ* hybridization zones as has been described for other species.[41–44]

We confirm previous findings that genetic structure in the giraffe is pronounced,[4,45] and although we found evidence for recent gene flow between sublineages, we did not detect contemporary gene flow between the major lineages *in situ*, despite close geographic proximity in parts of their range. This is in line with a recent study, also describing varying levels of Nubian ancestry in Reticulated giraffes in the Laikipia landscape

(Kenya), which were inferred to reflect historical rather than recent gene flow.[46] We speculate that this could be explained by a dynamic historical geographic distribution of giraffe populations in which populations were periodically isolated in allopatry by unsuitable habitats, rivers, or rift valleys and subsequently brought into secondary contact. This framework is referred to as mixing-isolation-mixing (MIM),[47] which is also generally consistent with phylogeographical refugial theory.[48–50] Under this framework, also suggested by Brown et al.,[45] recurrent Pleistocene climate oscillations and the associated rainfall and vegetation changes[50–52] could be responsible for intermittent connectivity between giraffe populations. Since giraffes are highly dependent on open savannah and acacia for browsing,[53–56] repeated interglacial forest expansions during the late Pleistocene[50–52] would have severely restricted connectivity between populations. During these times, giraffe populations may have been pushed back into smaller pockets of suitable habitats, and temporarily small population sizes could have sped up differentiation as a result of genetic drift. In addition, various rivers and rift systems have been proposed to constitute barriers for giraffe movements.[2] Although the importance of specific riverine barriers would have varied over time with the distribution of rainfall across different catchment areas,[57] the genetic discontinuities we observe on either side of some river valleys (Figures 2B and 2C) suggest that they strongly limit gene flow in this species. Similar discontinuity across the Gregory rift, as well as previous observations of a phenotypically distinct giraffe population isolated within the Luangwa Valley,[4,58] and the absence of giraffe from the lower Zambezi valley,[59,60] despite highly suitable habitats, suggest that gene flow is also limited by extensive areas of rugged or broken terrain. This was particularly prominent in the case of the split between Masai and Masai-Selous in Tanzania, which was previously suggested to be caused by the Gregory rift.[61] Although other co-distributed species show similar patterns of pronounced genetic structure and deep intraspecific evolutionary splits,[48,49] the giraffe is rather extreme in this regard. Brown et al.[45] found considerable local genetic differentiation even between subpopulations 67–134 km apart in the Serengeti, Tanzania. The seasonality of rains may also affect the distribution of genetic structure between giraffe lineages in east Africa through differential timing of calving peaks.[62] The above may suggest an unusual propensity for genetic differentiation in giraffes at local scales, despite them being highly mobile browsers with home range sizes varying from 100[2,63–65] to 2,000 km.[2,66] Hence, cryptic—perhaps behavioral[67]—limitations to gene flow, even at modest geographical scales, could reinforce any genetic isolation brought about by phylogeographical barriers caused by varying paleoclimates and tectonic uplift. Therefore, pronounced genetic structure and high genetic differentiation need not be indicative of long-standing genetic isolation between giraffe populations or lineages but could instead reflect small genetic neighborhood sizes.

In the light of the differentiation between the four major lineages, we do encourage separate assessments and management plans to conserve the diversity of giraffes most effectively. We further emphasize that insights into the substructure within these lineages and the overall distribution of genetic diversity are crucial for effective conservation of biodiversity, as it can inform prioritization, translocations, and *ex situ* breeding programs. Whether these major lineages will be designated as separate species or regarded as "separately evolving lineages" in a future assessment should not influence the goal to maximize the retention of genetic diversity, across and within lineages in this charismatic taxon. Further research is warranted to assess the long-term fate of populations affected by natural or human-mediated gene flow through translocations and in particular whether individuals resulting from admixture between major giraffe lineages have reduced fitness. This study not only provides new insights into the intertwined nature of evolution for a charismatic mega-herbivore but also provides genomic resources with high relevance to conservation programs that aim to maintain genetic diversity across major lineages and sublineages of giraffes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Mapping and reference genome quality filtering
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sample quality filtering
  - Relatedness
  - SNPs, genotypes and genotype imputation
  - Population structure, admixture and homogeneity
  - F_{ST} and EEMS
  - Runs of homozygosity, heterozygosity and demographic history
  - D-statistics, f-branch statistics, QPgraph and fastsimcoal

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.02.051.

ACKNOWLEDGMENTS

We thank Amal Al-Chaer for her help with DNA extractions. We are indebted to Peter Arctander, who organized the collection of samples in the 1980s and 1990s, and to the game and wildlife management authorities in various African countries for granting permission to use tissue samples for scientific purposes. G.G.E. and R.H. were supported by a Danmarks Frie Forskningsfond Sapere Aude research grant (DFF8049-00098B), and R.H. is further supported by a Carlsberg Young Researcher grant (CF21-0497). A.A. and M.S.R are supported by the Independent Research Fund Denmark (grant number: 8021-00360B). A.A., R.F.B., Z.L., and L.L. are funded by the Novo Nordisk Foundation (NNF20OC0061343). A.B.O. is supported by a Carlsberg Foundation Reintegration Fellowship (CF19-0427). I.M., X.L., and R.F.B. are supported by a Villum Young Investigator grant (VIL19114) awarded to I.M. M.H.S.S. is supported by a Carlsberg Foundation Reintegration Fellowship (CF20-0355).

AUTHOR CONTRIBUTIONS

Methodology: L.D.B., L.Q., K.H., G.G.E., M.S.R., R.F.B., J.M., T.B., X.W., L.L., C.N., X.L., Z.L., A.B.O., J.K., M.S., and C.S. Data curation: L.D.B. Visualization: L.D.B. and L.Q. Writing - original draft: L.D.B., L.Q., K.H., I.M., A.A., and R.H. Writing - review & editing: M.C., Y.M., M.A., M.H.S.S., V.M., C.M., and H.R.S. Supervision: I.M., A.A., and R.H. All authors contributed to the writing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 11, 2023
Revised: September 29, 2023
Accepted: February 21, 2024
Published: March 12, 2024

REFERENCES

1. Bercovitch, F.B., Berry, P.S.M., Dagg, A., Deacon, F., Doherty, J.B., Lee, D.E., Mineur, F., Muller, Z., Ogden, R., Seymour, R., et al. (2017). How many species of giraffe are there? Curr. Biol. *27*, R136–R137.
2. Bercovitch, F.B. (2020). Giraffe taxonomy, geographic distribution and conservation. Afr. J. Ecol. *58*, 150–158.
3. Petzold, A., and Hassanin, A. (2020). A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla). PLoS One *15*, e0217956.
4. Coimbra, R.T.F., Winter, S., Kumar, V., Koepfli, K.-P., Gooley, R.M., Dobrynin, P., Fennessy, J., and Janke, A. (2021). Whole-genome analysis of giraffe supports four distinct species. Curr. Biol. *31*, 2929–2938.e5.
5. Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, M.A., Vamberger, M., Fritz, U., and Janke, A. (2016). Multi-locus analyses reveal four giraffe species instead of one. Curr. Biol. *26*, 2543–2549.
6. Winter, S., Fennessy, J., and Janke, A. (2018). Limited introgression supports division of giraffe into four species. Ecol. Evol. *8*, 10156–10165.
7. Mallet, J., Besansky, N., and Hahn, M.W. (2016). How reticulated are species? Bioessays *38*, 140–149.
8. Fontsere, C., de Manuel, M., Marques-Bonet, T., and Kuhlwilm, M. (2019). Admixture in mammals and how to understand its functional implications: On the abundance of gene flow in mammalian species, its impact on the genome, and roads into a functional understanding. Bioessays *41*, e1900123.
9. Gopalakrishnan, S., Sinding, M.-H.S., Ramos-Madrigal, J., Niemann, J., Samaniego Castruita, J.A., Vieira, F.G., Carøe, C., Montero, M.d.M., Kuderna, L., Serres, A., et al. (2018). Interspecific gene flow shaped the evolution of the genus *Canis*. Curr. Biol. *28*, 3441–3449.e5.
10. Li, G., Figueiró, H.V., Eizirik, E., and Murphy, W.J. (2019). Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. Mol. Biol. Evol. *36*, 2111–2126.
11. Figueiró, H.V., Li, G., Trindade, F.J., Assis, J., Pais, F., Fernandes, G., Santos, S.H.D., Hughes, G.M., Komissarov, A., Antunes, A., et al. (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. *3*, e1700299.
12. Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M.A., and Janke, A. (2017). The evolutionary history of bears is characterized by gene flow across species. Sci. Rep. *7*, 46487.
13. Muller, Z., Bercovitch, F., Brand, R., Brown, D., Brown, M., Bolger, D., Carter, K., Deacon, F., Doherty, J.B., Fennessy, J., et al. (2016). *Giraffa camelopardalis* (The IUCN Red List of Threatened Species).
14. Dagg, A.I. (1962). The subspeciation of the giraffe. J. Mammal. *43*, 550–552.
15. O'Connor, D., Stacy-Dawes, J., Muneza, A., Fennessy, J., Gobush, K., Chase, M.J., Brown, M.B., Bracis, C., Elkan, P., Zaberirou, A.R.M., et al. (2019). Updated geographic range maps for giraffe, *Giraffa* spp., throughout sub-Saharan Africa, and implications of changing distributions for conservation. Mamm Rev. *49*, 285–299.
16. Muller, Z., Bercovitch, F., Brand, R., Brown, D., Brown, M., and Bolger, D. (2018). *Giraffa camelopardalis* (amended version of 2016 assessment) (The IUCN Red List of Threatened Species).
17. Muller, Z., Lee, D.E., Scheijen, C.P.J., Strauss, M.K.L., Carter, K.D., and Deacon, F. (2020). Giraffe translocations: A review and discussion of considerations. Afr. J. Ecol. *58*, 159–171.
18. Bertola, L.D., Miller, S.M., Williams, V.L., Naude, V.N., Coals, P., Dures, S.G., Henschel, P., Chege, M., Sogbohossou, E.A., Ndiaye, A., et al. (2022). Genetic guidelines for translocations: Maintaining intraspecific diversity in the lion (*Panthera leo*). Evol. Appl. *15*, 22–39. https://doi.org/10.1111/eva.13318.
19. Bertola, L.D., Sogbohossou, E.A., Palma, L., and Miller, S.M. (2022). Policy implications from genetic guidelines for the translocations of lions. Cat. News *75*, 41–43.
20. Agaba, M., Ishengoma, E., Miller, W.C., McGrath, B.C., Hudson, C.N., Bedoya Reina, O.C., Ratan, A., Burhans, R., Chikhi, R., Medvedev, P., et al. (2016). Giraffe genome sequence reveals clues to its unique morphology and physiology. Nat. Commun. *7*, 11519.
21. Liu, C., Gao, J., Cui, X., Li, Z., Chen, L., Yuan, Y., Zhang, Y., Mei, L., Zhao, L., Cai, D., et al. (2021). A towering genome: Experimentally validated adaptations to high blood pressure and extreme stature in the giraffe. Sci. Adv. *7*, eabe9459. https://doi.org/10.1126/sciadv.abe9459.
22. Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., et al. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science *364*, eaav6202. https://doi.org/10.1126/science.aav6202.
23. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
24. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
25. Garcia-Erill, G., and Albrechtsen, A. (2020). Evaluation of model fit of inferred admixture proportions. Mol. Ecol. Resour. *20*, 936–949.
26. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
27. Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. (2013). Estimating and interpreting F_{ST}: the impact of rare variants. Genome Res. *23*, 1514–1521. https://doi.org/10.1101/gr.154831.113.
28. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100.
29. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.
30. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
31. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.
32. Sjödin, P., McKenna, J., and Jakobsson, M. (2021). Estimating divergence times from DNA sequences. Genetics *217*, iyab008. https://doi.org/10.1093/genetics/iyab008.
33. Excoffier, L., and Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics *27*, 1332–1334.
34. Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. *21*, 584–595.

35. Maier, R., Flegontov, P., Flegontova, O., Changmai, P., and Reich, D. (2022). On the limits of fitting complex models of population history to genetic data. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.08.491072.

36. Petersen, M., Winter, S., Coimbra, R., J de Jong, M., Kapitonov, V.V., and Nilsson, M.A. (2021). Population analysis of retrotransposons in giraffe genomes supports RTE decline and widespread LINE1 activity in Giraffidae. Mob. DNA *12*, 27.

37. Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366*, 594–599.

38. Mayr, E. (1963). Animal Species and Evolution (Harvard University Press).

39. Lackey, L.B. (2011). Giraffe studbook *Giraffa camelopardalis* North American regional/global (World Association of Zoos and Aquariums). (WAZA)).

40. KWS (2018). National Recovery and Action Plan for Giraffe (*Giraffa camelopardalis*) in Kenya, pp. 2018–2022.

41. Bertola, L.D., Vermaat, M., Lesilau, F., Chege, M., Tumenta, P.N., Sogbohossou, E.A., Schaap, O.D., Bauer, H., Patterson, B.D., White, P.A., et al. (2022). Whole genome sequencing and the application of a SNP panel reveal primary evolutionary lineages and genomic variation in the lion (*Panthera leo*). BMC Genom. *23*, 321.

42. Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R.E., Georgiadis, N.J., Roca, A.L., and Hofreiter, M. (2010). Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. PLoS Biol. *8*, e1000564.

43. Zinner, D., Groeneveld, L.F., Keller, C., and Roos, C. (2009). Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? BMC Evol. Biol. *9*, 83.

44. Smitz, N., Berthouly, C., Cornélis, D., Heller, R., van Hooft, P., Chardonnet, P., Caron, A., Prins, H., van Vuuren, B.J., De Iongh, H., and Michaux, J. (2013). Pan-African genetic structure in the African buffalo (*Syncerus caffer*): investigating intraspecific divergence. PLoS One *8*, e56235.

45. Brown, D.M., Brenneman, R.A., Koepfli, K.-P., Pollinger, J.P., Milá, B., Georgiadis, N.J., Louis, E.E., Jr., Grether, G.F., Jacobs, D.K., and Wayne, R.K. (2007). Extensive population genetic structure in the giraffe. BMC Biol. *5*, 57.

46. Coimbra, R.T.F., Winter, S., Muneza, A., Fennessy, S., Otiende, M., Mijele, D., Masiaine, S., Stacy-Dawes, J., Fennessy, J., and Janke, A. (2023). Genomic analysis reveals limited hybridization among three giraffe species in Kenya. BMC Biol. *21*, 215.

47. He, Z., Li, X., Yang, M., Wang, X., Zhong, C., Duke, N.C., Wu, C.-I., and Shi, S. (2019). Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. Natl. Sci. Rev. *6*, 275–288.

48. Lorenzen, E.D., Heller, R., and Siegismund, H.R. (2012). Comparative phylogeography of African savannah ungulates. Mol. Ecol. *21*, 3656–3670.

49. Bertola, L.D., Jongbloed, H., van der Gaag, K.J., de Knijff, P., Yamaguchi, N., Hooghiemstra, H., Bauer, H., Henschel, P., White, P.A., Driscoll, C.A., et al. (2016). Phylogeographic patterns in Africa and high Resolution delineation of genetic clades in the lion (*Panthera leo*). Sci. Rep. *6*, 30807.

50. Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. Lond. B Biol. Sci. *359*, 183–195.

51. De Vivo, M., and Carmignotto, A.P. (2004). Holocene vegetation change and the mammal faunas of South America and Africa. J. Biogeogr. *31*, 943–957.

52. Kingdon, J. (2018). The Kingdon Field Guide to African Mammals, Second Edition (Bloomsbury Publishing).

53. Bond, W.J., and Loffell, D. (2001). Introduction of giraffe changes acacia distribution in a South African savanna. Afr. J. Ecol. *39*, 286–294.

54. Ciofolo, I., and Le Pendu, Y. (2002). The feeding behaviour of Giraffe in Niger. Mammalia *66*, 183–194.

55. Berry, P.S.M., and Bercovitch, F.B. (2017). Seasonal and geographical influences on the feeding ecology of giraffes in the Luangwa Valley, Zambia: 1973-2014. Afr. J. Ecol. *55*, 80–90. https://doi.org/10.1111/aje.12324.

56. Pellew, R.A. (1984). The feeding ecology of a selective browser, the giraffe (*Giraffa camelopardalis tippelskirchi*). J. Zool. *202*, 57–81.

57. Kingdon, J. (2013). Mammals of Africa (Bloomsbury).

58. Fennessy, J., Bock, F., Tutchings, A., Brenneman, R., and Janke, A. (2013). Mitochondrial DNA analyses show that Zambia's South Luangwa Valley giraffe (*Giraffa camelopardalis thornicrofti*) are genetically isolated. Afr. J. Ecol. *51*, 635–640.

59. Dunham, K.M., Mackie, C.S., Nyaguse, G., and Zhuwau, C. (2014). Aerial survey of elephants and other large herbivores in the Sebungwe (Zimbabwe). https://zamsoc.org/s/Great-Elephant-Census-Sebungwe-Survey-Report-2015.pdf.

60. Fennessy, J., Fennessy, S., Brown, M.B., Ferguson, S., Muneza, A., Bocchino, C., Nyambe, N., and Bollmann, N. (2022). Kavango-Zambezi Transfrontier Conservation Area Giraffe Conservation Strategy 2022, p. 26.

61. Lohay, G.G., Lee, D.E., Wu-Cavener, L., Pearce, D.L., Hou, X., Bond, M.L., and Cavener, D.R. (2023). Genetic evidence of population subdivision among Masai giraffes separated by the Gregory Rift Valley in Tanzania. Ecol. Evol. *13*, e10160.

62. Thomassen, H.A., Freedman, A.H., Brown, D.M., Buermann, W., and Jacobs, D.K. (2013). Regional differences in seasonal timing of rainfall discriminate between genetically distinct East African giraffe taxa. PLoS One *8*, e77191.

63. Knüsel, M.A., Lee, D.E., König, B., and Bond, M.L. (2019). Correlates of home range sizes of giraffes, *Giraffa camelopardalis*. Anim. Behav. *149*, 143–151.

64. McQualter, K.N., Chase, M.J., Fennessy, J.T., McLeod, S.R., and Leggett, K.E.A. (2016). Home ranges, seasonal ranges and daily movements of giraffe (*Giraffa camelopardalis giraffa*) in northern Botswana. Afr. J. Ecol. *54*, 99–102.

65. Flanagan, S.E., Brown, M.B., Fennessy, J., and Bolger, D.T. (2016). Use of home range behaviour to assess establishment in translocated giraffes. Afr. J. Ecol. *54*, 365–374.

66. Fennessy, J. (2009). Home range and seasonal movements of *Giraffa camelopardalis angolensis* in the northern Namib Desert. Afr. J. Ecol. *47*, 318–327.

67. Pratt, D.M., and Anderson, V.H. (2010). Giraffe cow-calf relationships and social development of the calf in the Serengeti. Z. Tierpsychol. *51*, 233–251. https://doi.org/10.1111/j.1439-0310.1979.tb00686.x.

68. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082.

69. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

70. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

71. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

72. Smit, A.F.A., Hubley, R., and Green, P. (2004). Repeat-Masker Open-3.0. http://www.repeatmasker.org.

73. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinf. *15*, 356.

74. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.

75. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

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

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

78. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., and Others. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

79. Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinf. *12*, 246.

80. Moltke, I., and Albrechtsen, A. (2014). RelateAdmix: a software tool for estimating relatedness between admixed individuals. Bioinformatics *30*, 1027–1028.

81. Browning, B.L., and Browning, S.R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. *84*, 210–223.

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

83. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

84. Köster, J., and Rahmann, S. (2018). Snakemake-a scalable bioinformatics workflow engine. Bioinformatics *34*, 3600.

85. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.

86. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

87. Rogers, A.R. (2014). How population growth affects linkage disequilibrium. Genetics *197*, 1329–1341.

88. Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genom. *21*, 94.

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

90. Suraud, J.-P., Fennessy, J., Bonnaud, E., Issa, A.M., Fritz, H., and Gaillard, J.-M. (2012). Higher than expected growth rate of the endangered West African giraffe *Giraffa camelopardalis peralta*: a successful human–wildlife cohabitation. Oryx *46*, 577–583.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69506
Deposited data		
Raw sequencing reads	This study	ENA accession numbers: PRJEB66216
Software and algorithms		
PALEOMIX	Schubert et al.[68]	https://github.com/MikkelSchubert/paleomix/tree/pub/2022/africa
AdapterRemoval v2.3.3	Schubert et al.[69]	https://github.com/MikkelSchubert/adapterremoval
BWA mem v0.7.17-r1188	Li et al.[70]	http://bio-bwa.sourceforge.net/
Samtools v1.11	Li et al.[71]	http://www.htslib.org/
RepeatMasker v.4.1.1	Smit et al.[72]	http://www.repeatmasker.org/
ANGSD	Korneliussen et al.[73]	http://www.popgen.dk/angsd/index.php/ANGSD
PCAngsd	Meisner et al.[74]	http://www.popgen.dk/software/index.php/PCAngsd
FastQC	Andrews et al.[75]	https://github.com/s-andrews/FastQC
MultiQC	Ewels et al.[76]	https://multiqc.info/
realSFS	Nielsen et al.[77]	http://www.popgen.dk/angsd/index.php/RealSFS
PLINK v1.9	Chang et al.[78]	https://www.cog-genomics.org/plink/
BCFTools v1.9 and v1.16	Li et al.[30]	https://samtools.github.io/bcftools/
ADMIXTURE	Alexander et al.[79]	https://dalexander.github.io/admixture/
EvalAdmix	Garcia-Erill et al.[25]	https://github.com/GenisGE/evalAdmix
relateAdmix	Moltke et al.[80]	http://www.popgen.dk/software/index.php/RelateAdmix
BEAGLE v3.3.2	Browning et al.[81]	http://faculty.washington.edu/browning/beagle/b3.html
AdmixTools	Patterson et al.[26]	https://github.com/DReichLab/AdmixTools
RStudio v2022.07.2	RStudio Team	https://rstudio.com/
LDdecay	In house script	https://github.com/aalbrechtsen/LDdecay
EEMS	Petkova et al.[28]	https://github.com/dipetkov/eems
R package coda	Plummer et al.[82]	https://cran.r-project.org/package=coda
PSMC	Li et al.[30]	https://github.com/lh3/psmc
Dsuite	Malinsky et al.[34]	https://github.com/millanek/Dsuite
R package ape	Paradis et al.[83]	https://cran.r-project.org/package=ape
Two-Two (TT) method	Sjödin et al.[32]	https://github.com/jammc313/TT-method
AdmixTools2	Browning et al.[81]	https://github.com/uqrmaie1/admixtools
Fastsimcoal2 (fsc27)	Excoffier et al.[33]	http://cmpg.unibe.ch/software/fastsimcoal27/
Snakemake	Köster et al.[84]	https://snakemake.readthedocs.io/en/stable/
Other		
Giraffe WGS data	Coimbra et al.[4]	NCBI BioProject: PRJNA635165
Giraffe WGS data	Agaba et al.[20]	GenBank: SRR3218456
Giraffe reference genome	Liu et al.[21]	GenBank: GCA_017591445.1
Okapi reference genome	Chen et al.[22]	Ruminant Genome Database: http://animal.omics.pro/code/index.php/Ruminantia
Okapi outgroup	Agaba et al.[20]	GenBank: SRR3217625
Pronghorn outgroup	Chen et al.[22]	Ruminant Genome Database: http://animal.omics.pro/code/index.php/Ruminantia

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rasmus Heller (rheller@bio.ku.dk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data generated for this study have been uploaded to the European Nucleotide Archive and are publicly available as of the date of publication under accession number ENA: PRJEB66216, with samples having accession numbers ERS16390370-ERS16390416.
- The code used for analyses in this project is available on the project's github: https://github.com/popgenDK/seqAfrica_giraffes/ . Some pipelines were developed using Snakemake.[84]
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

A total of 47 giraffe samples (Table S1) from the existing collection at Copenhagen University, covering all four major lineages were selected for genome sequencing. Samples were selected based on reliable information regarding their sampling locality, and based on the aim of filling gaps in existing giraffe phylogeographic datasets. A particular focus was placed on the East African region, where the ranges of three major giraffe lineages connect. All samples were collected between 1993 and 2000, in collaboration with local wildlife authorities and in compliance with local and international legislation. Samples were stored in DMSO in -80°C until DNA extraction.

METHOD DETAILS

For DNA extraction, we used the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA, USA), following the manufacturer's protocol. Subsequently, RNase was added to the samples to ensure they consist of RNA-free genomic DNA. DNA concentrations were then measured with a Qubit 2.0 Fluorometer and a Nanodrop before using gel electrophoresis to check the quality of the genomic DNA. Library preparation and DNA sequencing was conducted at GrandOmics, China. Briefly, genomic DNA from each sample was randomly fragmented, and genomic DNA was size selected to around 330-530bp insert size. The fragments were subjected to end-repair and then 3′ adenylated. Adaptors were ligated to the ends of these 3′ adenylated fragments, and PCR was performed to amplify the product. Next, single-stranded PCR products were produced via denaturation, and circularized. Single-stranded circular products were used, while non-circular (linear) DNA molecules were digested. DNA concentration and library fragment lengths were checked before pooling libraries for sequencing. DNA sequencing was done using MGISEQ-T7 technology, with 150 bp paired-end reads. To build upon existing datasets, we downloaded whole-genomes sequencing datasets for 43 giraffes with known origin (Table S1) from the National Center for Biotechnology Information's Short Read Archive.[4,20] In addition, we downloaded a giraffe[21] and an okapi genome[20] to be used as references for mapping, and a second okapi[22] genome as an outgroup and a pronghorn[22] genome for polarization (Table S1).

Mapping and reference genome quality filtering

Sequencing reads from all 90 giraffe samples were mapped to two reference genomes, giraffe ([21], GenBank: GCA_017591445.1) and okapi ([22], Ruminant Genome Database: http://animal.omics.pro/code/index.php/RGD) (Table S1), using the development version of the Binary Alignment Map (BAM) pipeline included with PALEOMIX[68]; git branch 'pub/2022/africa'). This is a pipeline designed for processing demultiplexed high-throughput sequencing data. Before mapping, adapter sequences were trimmed using AdapterRemoval v2.3.3,[69] using Universal Illumina adapter sequences for the Illumina data (previously published giraffe genomes) and recommended adapter sequences for the genomes generated from our samples. Overlapping read pairs were merged to increase the fidelity of base calls and to increase specificity in the mapping procedure. Merging was done with the –collapse-conservatively option, which assigns 'N' to any overlapping position with contradictory base calls with equal quality scores. Empty (zero length) reads resulting from primer-dimers were discarded. No trimming of ambiguous and low quality bases was performed.

Trimmed paired and merged reads were mapped using BWA mem v0.7.17-r1188,[70] duplicates reads were marked using samtools 'markdup' v1.11[71] (for paired reads) and PALEOMIX 'rmdup_collpased' (for merged reads). The resulting BAMs were filtered to remove unmapped reads and reads with unmapped mates, secondary alignments, QC-failed reads, duplicate reads, and supplementary alignments. In addition, we removed alignments with an inferred insert size less than 50 or greater than 1000 bp, and reads where

fewer than 50 bp or less than 50% bases were mapped to the reference, and we removed paired reads mapping to different contigs or in an unexpected orientation.

We identified problematic regions of the giraffe and okapi reference genomes and discarded reads mapping to those regions prior to analysis. First, we used RepeatMasker v.4.1.1 to exclude low complexity and repeat regions with the engine of rmblast, and commands '-frag 50000 -species Mammalia'.[72] Based on per-site global sequencing depth distribution calculated using ANGSD[73] across all mapped samples, we excluded sites with a sequencing depth in the lowest or highest 1% of the distribution. We used a heterozygosity filter to prevent the emergence of sites with excess heterozygosity, likely reflecting low genome assembly quality or paralog regions. To identify these regions, we used PCAngsd[74] in which we calculated per site inbreeding coefficients (F) in a way that accounts for population structure and missing genotypes.[74] The first three principal components were used to account for population structure when obtaining individual allele frequencies. We excluded 100kb windows containing sites that were found to have excessive heterozygosity ($F < -0.95$, p value<1e-10).

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample quality filtering

We assessed the quality of the mapping using FastQC,[75] MultiQC[76] and ANGSD.[73] We identified samples with high error rates by following the 'perfect-individual' approach from Orlando et al. (2013)[85] as implemented in ANGSD.[73] Here, excess mismatches between each sample and the outgroup are compared to the mismatches between a consensus sequence of a high depth sample (the 'perfect individual') and the outgroup. To avoid reference bias and ensure equal genetic distance between the giraffe samples and the outgroup, okapi was selected as the outgroup. Assuming a similar number of derived alleles within the ingroup, an excess of derived alleles in a specific individual suggests a high error rate. Based on sequencing depth and overall mapping statistics we selected sample GCamTzC_9108 to generate a consensus sequence, using -doCounts 1 -doFasta 2 in ANGSD. Error rates were estimated using -doAncError 2 in ANGSD. Data were restricted to a minimum base quality of 30 and a minimum mapping quality of 20, both for generating the consensus and for the error estimation.

Relatedness

To avoid bias from the inclusion of closely related individuals, we identified first degree relatives based on kinship coefficient directly from the aligned sequencing data. This was done by calculating the genotype likelihood in ANGSD based on the GATK model and then calculating the two dimensional SFS for each pair of individuals using realSFS.[77] From the SFS, we calculated the kinship coefficient using the KING estimator.[86] Duplicates and/or monozygotic twins have an expected kinship coefficient of ≈ 0.5 and first degree related individuals have an expected coefficient of ≈ 0.25. The kinship coefficient can be further biased due to population structure, and therefore we performed additional analyses to identify potential second degree relatives. From the imputed genotype calls (see below: SNPs, genotypes and genotype imputation), we assessed homogeneity within sublineages based on PCA from PLINK v1.9[24] and admixture proportions from ADMIXTURE,[23] with evaluation from EvalAdmix[25] (see below: Admixture). We inferred the fraction of Identity-By-Descent (IBD) sharing between individuals within homogeneous sublineages, using the moment estimator implemented in PLINK. We performed the analysis based on SNPs with allele frequencies above 0.05 and missingness below 0.05, based on within-sublineage estimates. Simultaneously, we identified admixed individuals that have more than 10% admixture proportions from two or more ancestries and inferred the IBD sharing by relateAdmix.[80] We considered second degree relatives as having a fraction of loci sharing one allele IBD above 0.375.

SNPs, genotypes and genotype imputation

For downstream analyses, we generated three datasets, tailored to the specific demands of various methods: 1) all SNPs: all variable, diallelic positions, masking genotypes with depth less than 10 as well as heterozygous genotypes supported by less than two reads per allele (43,723,187 sites), 2) all genotypes: called for samples individually, all positions, except multiallelic sites and indels, but including monomorphic positions, and with the same genotype masking as for all SNPs, and 3) imputed genotypes: genotypes were determined based on the imputation results and the allelic R2 file outputted by BEAGLE v3.3.2,[81] and sites with R2<0.99 and MAF<0.05 were removed for downstream analyses (12,187,185 sites). In all three cases, the reference genome quality filters were applied as described above.

SFS estimation was based on all genotypes (dataset 2). For the joint SFS of a particular pair of sublineages, we then filtered sites with any missing genotypes in the relevant sublineages, and created the SFS by counting the observed allele count combinations.

Population structure, admixture and homogeneity

Population structure was assessed for 78 individuals, i.e., excluding samples with quality issues and one closely related individual (see Results), using the imputed genotypes (dataset 3). We estimated admixture proportions using ADMIXTURE.[79] ADMIXTURE was run from $K = 2$ to $K = 10$ with 100 independent runs for each K, using different seeds, where the runs with the best log-likelihoods were chosen. We then used EvalAdmix[25] to estimate the pairwise correlation of residuals between the observed genotypes and the genotypes predicted by the estimated allele frequencies and admixture proportions. The correlation can be used as a measure of model fit, with positive correlation of residuals between samples indicating relatedness or a pair from the same population without its own ancestry component. Likewise, a negative correlation indicates that the individuals should not be in the same ancestry

component. At K = 9 a cluster emerged within Reticulated giraffes which was the result of previously undetected family structure, and is in line with previously reported results for these individuals.[4] Since this structure does not reflect a meaningful cluster from the evolutionary perspective, we excluded the four individuals forming the additional cluster (GCamKeC_7734, GCamKeC_7727, GCamKeC_7725, GCamKeC_7733) to explore further underlying structure, leading to a split occurring within Masai giraffes (Figure S1). For downstream analysis, we treat Reticulated giraffes, including the related individuals, as a single group, while we also use the split between Masai and Masai-Selous, leading to a total of nine sublineages. Five individuals were flagged as admixed, as they showed admixture proportions of >10%, and were excluded from some downstream analyses, if appropriate. In addition, we performed PCA using PLINK v1.9[78] on the VCF, which was also used for the ADMIXTURE analysis (dataset 3). We explored patterns of differentiation across the first ten principal components.

To assess the homogeneity of each of the nine sublineages as inferred by ADMIXTURE, we applied D-statistics (ABBA-BABA) based on all SNPs (dataset 1), as implemented by the function qpDstat from AdmixTools.[26] We tested each pair of individuals within a sublineage (H1,H2) against an individual from another sublineage within the same major lineage (and an individual from Nubian for Reticulated giraffe) (H3) and the okapi as an outgroup, using a block size of 5 million bp. Based on these results, we excluded two individuals (GCamTzW_3369 Masai, GCamZwS_1546 Southern African) for downstream gene flow estimates and inferring relationships between sublineages.

Recent admixture can confound estimates of ancient gene flow and divergence times, illustrating the importance of testing for homogeneity and recent admixture in the units used for downstream analyses. We used LD decay curves to assess recent admixture[87] of the nine identified giraffe sublineages and therefore their suitability for evolutionary modeling. We generated LDdecay curves using R-scripts from https://github.com/aalbrechtsen/LDdecay and data from a single chromosome. Chromosome 1 was selected for the analysis as it is one of the large chromosomes (>200Mb) in the giraffe genome assembly. We used a sliding window of 800 bp which represents the number of SNPs on each side of the focal SNP and bin sizes of 0.1 Mbp. Adequate depth for our choice of binning was ensured by using R-scripts from https://github.com/aalbrechtsen/LDdecay, which showed sufficient data for analysis using the parameter values selected above. Because the sample size varied between the sublineages (from N = 4 to N = 12 individuals) we downsampled the number of individuals per sublineage to four individuals (Table S1) when performing the analysis to allow between lineage comparisons. See also Figure S2.

F_{ST} and EEMS

For each pair of major lineages and each pair of sublineages, we calculated F_{ST} using Hudson's estimator[27] based on the two-dimensional SFS. Connectivity on a continental scale was explored by running EEMS,[28] which assesses the decay of genetic similarity between sampling locations under a null model of isolation by distance. We created an average genetic dissimilarity matrix based on imputed genotypes (dataset 3), using the bed2diffs function as is implemented in EEMS.[28] For calculating the geographic distance between sampling locations, we used the geodist() function in R. EEMS was run over 400 demes, for 40 million generations, discarding the first 20 million as burnin. Convergence of three independent runs was assessed visually and by using the Gelman–Rubin diagnostic in the R package coda.[82] The same approach was repeated on a smaller geographic scale, to infer patterns of reduced gene flow within East Africa.

Runs of homozygosity, heterozygosity and demographic history

ROH were inferred based on all SNPs (dataset 1), using the command '–homozyg' in PLINK v1.9[78] with default settings, except allowing a maximum of 3 heterozygous calls and 20 missing calls when detecting ROH (–homozyg-window-het 3 –homozyg-window-missing 20). Minimum ROH length was set to the default of 1 Mbp. Fractions of the genome as ROH (FROH) were calculated by dividing the total length of ROH per sample by the entire genome length. Using a published approach to estimate the entire genome length per sample,[88] 'pseudo-homozygous' individuals were created with each individual's genotype fixed to be homozygous for all sites. The same PLINK ROH command was used and the total length of ROH from these individuals was taken as the total genome length.

Genome-wide heterozygosity was calculated for each sample based on called genotypes and additionally adjusted taking these extended homozygous portions of the genome into account. This ROH adjusted heterozygosity was calculated as: ROH adjusted heterozygosity = heterozygosity / (1 - FROH). Both results are shown in Figure 3A.

We used PSMC[30,89] to infer changes in historical population size based on all genotypes (dataset 2). PSMC was run on all individuals except for those which were flagged as admixed or deviating from homogeneity. We used default settings, a mutation rate of 2.1×10^{-8} per generation[22] and generation time of 10 years.[90] Additionally, to explore effects of recent inbreeding on PSMC inferences, we also ran PSMC only based on the genomic regions excluding ROH. For each sample, we masked its ROHs identified by PLINK. Results for individual populations are shown in Data S1C.

D-statistics, f-branch statistics, QPgraph and fastsimcoal

We used the Dsuite package to summarize results from D-statistics for combinations of sublineages that logically fit a pre-specified phylogenetic tree.[34] The phylogenetic tree was based upon the topology of a neighbor-joining tree we inferred based on pairwise Identity-by-State (IBS) values between individuals, with okapi acting as an outgroup. In PLINK we used the commands '–distance

square ibs' and 'allele-ct' to estimate pairwise IBS values. We then used the ape plotting package in R to create an unrooted neighbor-joining tree with the 'nj' command,[83] excluding the five individuals which were flagged as recently admixed and two individuals deviating in the homogeneity test.

To estimate divergence times for the various splits in the tree, we applied the 'TT' (TT) method,[32] which is known to be robust for gene flow between populations. We selected one sample for each of the nine sublineages with similar coverage (≈17-19X; Table S1) to account for potential mapping biases, performed a separate variant calling and polarized against okapi and pronghorn as outgroups. We used the same mutation rate and generation length settings as for PSMC.

We then used the Dtrios function in Dsuite to calculate D-statistics for all possible trio combinations, followed by applying the Fbranch command of Dsuite to these D-statistics in order to calculate the f-branch statistics (a statistic related to the f-4 statistic). A heatmap was created to summarize the f-branch results within the suggested phylogenetic framework, which is provided in Newick tree format.

Following our results, indicating the role of substantial introgression between major giraffe lineages, we used admixture graphs to obtain a more detailed insight into the complex evolutionary history of giraffes. In contrast to tree-based approaches, admixture graphs allow for ancestral admixture events between lineages. Based on the Dsuite results, many of the violations of treeness involved the Nubian and Angolan sublineages, leading to an inflated number of admixture events within major lineages. In order to focus on introgression between major lineages, we therefore first explored models without these two sublineages, having seven sublineages in total.[35] We then subsequently explored the more complex model, including all 9 sublineages.

To explore plausible admixture graphs, we used the 'find_graphs' function in the AdmixTools2 package.[35] F2-scores were first extracted from a PLINK file using the 'extract_f2' function using a block length of 4×10^6 bp. Only SNPs with a maximal missingness of 0.9 within any given sublineages were retained. For each allowed number of admixture events (0-5), we ran 'find_graphs' 500 times. This function obtains a test score by optimizing a topology from a subset of sites and uses the remaining sites for testing as described in AdmixTools2.[35] Briefly, AdmixTools2 uses jackknife approach for significant testing where each of the highest scoring graphs are compared against all of the other graphs. If the comparison leads to a p value below 0.05, we interpret this as the highest scores graph being significantly better than the other graphs. We increased the number of allowed admixture events in the graph, starting from zero, until the best test score was not significantly better than the all of the best scoring graphs with fewer admixture events.

Based on the inferred topology from QPgraph, we used fastsimcoal to further investigate the inferred admixture events. However, fastsimcoal cannot deal with the 7 populations because it only allows for a limited number of entries in the multidimensional joint SFS (maximum of 1 million entries). Therefore we restricted the analysis to a subset of the populations most relevant for the inferred gene flow. We explored two different demographic models from the multidimensional joint SFS, excluding and including the Masai major lineage, shown in Figures S4A and S4B. The SFS was polarized to both okapi and pronghorn, and sites with missing data were removed for each set of populations separately. We also removed sites in which any of the outgroups were polymorphic, since we cannot polarize these sites with high certainty. Fastsimcoal was run 50 times for each model with 1 million simulations and 100 ECM cycles. The parameter estimates from the top 50% of the runs in terms of estimated log likelihood all gave consistent results and are shown in Figures S4C and S4D.