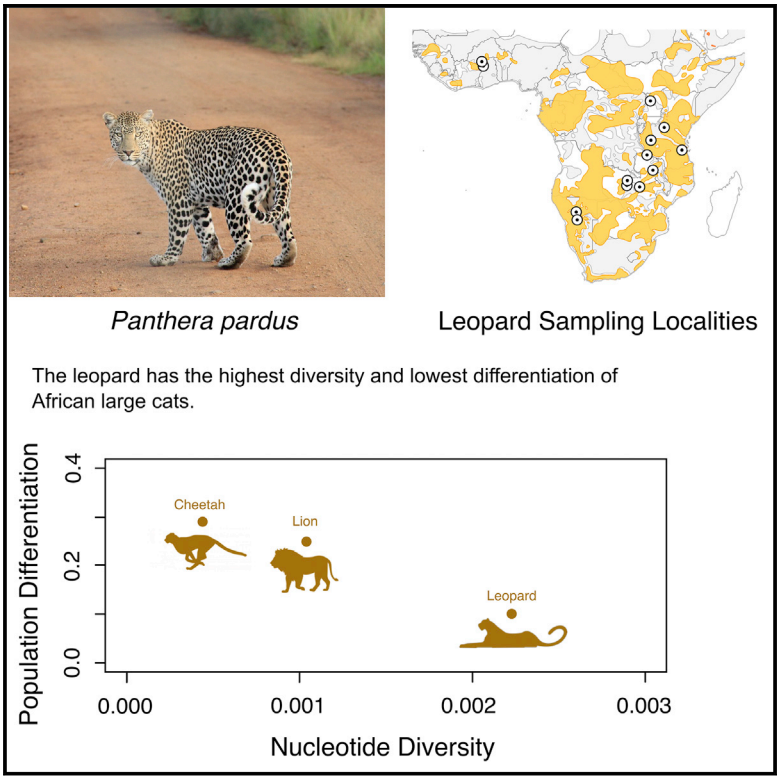
Article

Current Biology

High genetic diversity and low differentiation reflect the ecological versatility of the African leopard

Graphical Abstract



Highlights

- African and Amur leopards have markedly different demographic trajectories
- Among big cats, African leopards have the highest genetic diversity
- Gene flow on a continent-wide scale maintains low genetic differentiation
- Broad dietary and habitat niche likely explain the extraordinary genetic makeup

Authors

Patrícia Pečnerová, Genís Garcia-Erill, Xiaodong Liu, ..., Rasmus Heller, Ida Moltke, Kristian Hanghøj

Correspondence

rheller@bio.ku.dk (R.H.), ida@binf.ku.dk (I.M.), kristianhanghoej@gmail.com (K.H.)

In Brief

Pečnerová et al. analyze whole-genome data from 53 African leopards and identify genetic features that set the African leopard apart from the Amur leopards, as well as the other big cats. This includes long-term high effective population size and exceptionally high levels of genetic diversity and connectivity.

Pečnerová et al., 2021, Current Biology *31*, 1862–1871
May 10, 2021
https://doi.org/10.1016/j.cub.2021.01.064

CellPress

CellPress

Current Biology

Article

High genetic diversity and low differentiation reflect the ecological versatility of the African leopard

Patrícia Pečnerová,[1,5,7] Genís Garcia-Erill,[1,5] Xiaodong Liu,[1,5] Casia Nursyifa,[1,5] Ryan K. Waples,[1,5] Cindy G. Santander,[1,5] Liam Quinn,[1] Peter Frandsen,[1,2] Jonas Meisner,[1] Frederik Filip Stæger,[1] Malthe Sebro Rasmussen,[1] Anna Brüniche-Olsen,[1,3] Christian Hviid Friis Jørgensen,[1] Rute R. da Fonseca,[4] Hans R. Siegismund,[1] Anders Albrechtsen,[1,6] Rasmus Heller,[1,6,*] Ida Moltke,[1,6,*] and Kristian Hanghøj[1,6,8,*]

[1]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark
[2]Copenhagen Zoo, Research and Conservation, Roskildevej 32, 2000 Frederiksberg, Denmark
[3]Department of Forestry and Natural Resources, Purdue University, 610 Purdue Mall, West Lafayette, IN 47907, USA
[4]Center for Macroecology, Evolution and Climate (CMEC), GLOBE Institute, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
[5]These authors contributed equally
[6]Senior authors
[7]Twitter: @PatriciaChrzan
[8]Lead Contact
*Correspondence: rheller@bio.ku.dk (R.H.), ida@binf.ku.dk (I.M.), kristianhanghoej@gmail.com (K.H.)
https://doi.org/10.1016/j.cub.2021.01.064

SUMMARY

Large carnivores are generally sensitive to ecosystem changes because their specialized diet and position at the top of the trophic pyramid is associated with small population sizes. Accordingly, low genetic diversity at the whole-genome level has been reported for all big cat species, including the widely distributed leopard. However, all previous whole-genome analyses of leopards are based on the Far Eastern Amur leopards that live at the extremity of the species' distribution and therefore are not necessarily representative of the whole species. We sequenced 53 whole genomes of African leopards. Strikingly, we found that the genomic diversity in the African leopard is 2- to 5-fold higher than in other big cats, including the Amur leopard, likely because of an exceptionally high effective population size maintained by the African leopard throughout the Pleistocene. Furthermore, we detected ongoing gene flow and very low population differentiation within African leopards compared with those of other big cats. We corroborated this by showing a complete absence of an otherwise ubiquitous equatorial forest barrier to gene flow. This sets the leopard apart from most other widely distributed large African mammals, including lions. These results revise our understanding of trophic sensitivity and highlight the remarkable resilience of the African leopard, likely because of its extraordinary habitat versatility and broad dietary niche.

INTRODUCTION

Apex predators are more sensitive to climate change and other ecological disturbances than species at lower trophic levels.[1–3] This makes them more prone to demographic fluctuations and thus more susceptible to population crashes and even extinction.[4] Given that genetic diversity is linked to demographic fluctuations, apex predators are expected to have on average lower genetic diversity than their herbivorous prey species. A rapidly accumulating body of evidence from whole-genome studies[5,6] has confirmed this in, among others, all of the big cats, i.e., the species belonging to *Panthera*,[7] a genus of typical apex predators.

Among the big cats, the leopard, *Panthera pardus*, stands out because of its extensive geographic distribution[8] and its ability to persist under a wide range of ecological conditions, even close to human settlements, labeling it as a habitat and dietary generalist.[8,9] Given that genetic diversity is related to ecological resilience, the leopard should have higher genetic diversity than other big cats, yet this was not found in previous genome-wide studies.[5] However, the only published whole-genome data from the leopard[5,10] originates from the Amur leopard (*P. p. orientalis*), which lives at the eastern extremity of the species' geographic distribution (Figure S1). This population has a history of severe range and population contractions, making it the most critically endangered leopard subspecies with less than 60 individuals surviving in the wild.[11]

The leopard is thought to have originated in eastern Africa approximately 2–4 million years ago (mya).[12,13] The divergence between African and Asian leopards has been dated to ~710 thousand years ago (kya) on the basis of a mitogenome phylogeny calibrated by using historical and ancient samples.[12]

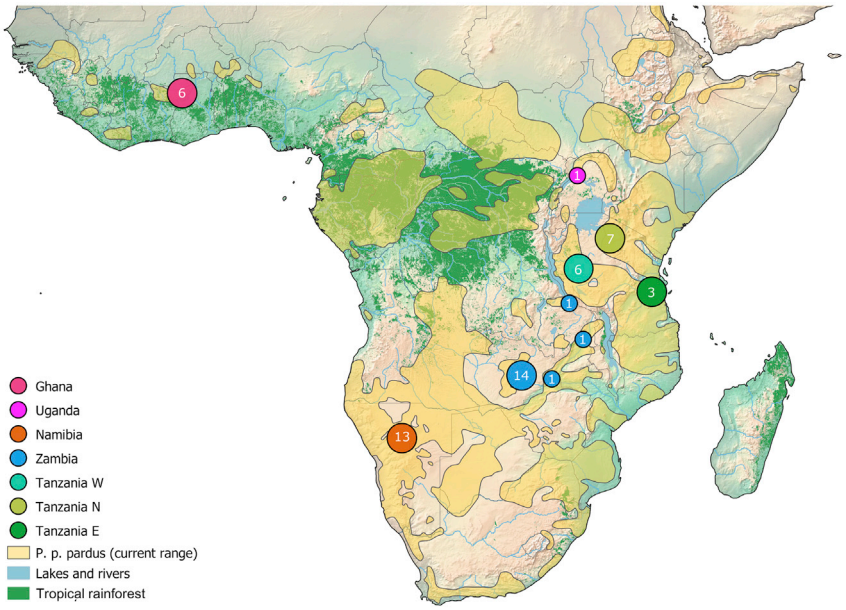


Figure 1. Sampling locations

Locations of origin of the samples analyzed in this study. Terrain and country borders are shown for context. Sample sizes at each location are shown within the colored circles. Information on the current range of *P. p. pardus* was drawn from the IUCN Red List of Threatened Species 2020.[11] The tropical rainforest is highlighted based on the extent inferred in Mayaux et al.[18] See also Figure S1 and Table S1.

Subsequently, the divergence between Asian and Pleistocene-era European leopards was dated to ~483 kya, and if we assume that this divergence reflects the out-of-Africa event, this sets the likely timing of leopards entering Eurasia to 710–483 kya.[12] Because of this long period of separation, Asian and African leopards might differ markedly in their evolutionary trajectories, including effective population size and population connectivity. This, combined with the Amur leopard's very limited census population size and the fact that population contractions reduce genetic diversity, suggests that the results of genomic analyses of Amur leopards might not represent the entire leopard species and in particular not the African leopards.

Although a few genetic studies have been performed on the African leopard, they have all been based on microsatellites and/or mitochondrial data. Based on these studies, which identified low population differentiation, all African leopards have been classified as a single subspecies *P. p. pardus*.[14–16] On the contrary, a recent fine-scale sampling of 182 African leopards by using 611 base pairs of mitochondrial DNA suggested substantial population structure and deep divergence within Africa.[17] Therefore, our genetic understanding of leopards in their proposed continent of origin remains conflicted and scarce.

In this study, we investigated population structure, demographic history, and genetic diversity in African leopards by sequencing and analyzing 53 whole genomes, covering most of their current range in Africa. The aim was to learn more about the evolutionary dynamics of the leopard in its biogeographic cradle and to assess whether the generalist ecology of the leopard sets it apart from other big cats.

RESULTS

We generated whole-genome sequencing data for 53 African leopards from 10 locations in sub-Saharan Africa and a total of 5 countries: Ghana, Namibia, Tanzania, Uganda, and Zambia (Figure 1; Table S1). We targeted 2–5× depth of coverage for 47 samples and deeply sequenced (to 15–20× depth of coverage) the remaining 6 samples. Mapping and rigorous quality filtering of the reference genomes and the samples (see STAR methods and Tables S2 and S3) resulted in a final dataset of 1,374,856,842 genomic sites when mapped to the leopard reference genome and 916,801,099 sites for the domestic cat (*Felis catus*) reference for 41 samples (36 low coverage and 5 high coverage). The 12 discarded samples were removed due to either low DNA quality, species mis-labeling, or sample duplication (Tables S2–S5). Our final dataset contained two pairs of first-degree relatives (Table S5), of which one individual per pair was removed for analyses where relatedness can confound the results. For all analyses including any low-depth samples, we accounted for the genotype uncertainty by using methods based on genotype likelihoods or single-read sampling instead of called genotypes (see STAR methods).

Population structure within African leopards

We first explored the population structure of African leopards by performing a principal component analysis (PCA) of 39 African leopards (excluding 1 sample from each of the 2 pairs of first-degree relatives) by using PCAngsd.[19] We grouped samples by country of origin, except for Tanzania, which we divided into three groups (North, West, and East). The resulting plot roughly mirrors the geography (Figure 2A). Specifically, the first principal component (PC) captures a cline of genetic variation from the north (Ghana and Tanzania North) to the southwest (Namibia), whereas the second PC captures a cline across the northern locations. Notably, sampling locations are not discretely clustered along these PC axes; instead, we detected a pattern consistent with continuous genetic variation across the populations, with the exception of the Ghana population.

Next, we estimated per-individual admixture proportions by using NGSadmix.[20] We excluded the single sample from Uganda (n = 38) because an imbalanced sample size can create biases in admixture models.[24] We found that a model with four ancestral source populations (K = 4) provides the best fit to the dataset (Figures 2B and S2). In this model, most sampling locations are assigned to homogeneous ancestry groups with two exceptions: first, Zambia is modeled as a mixture of the Namibia and Tanzania West clusters; and second, Tanzania West and East are grouped together, which might be caused by a small sample

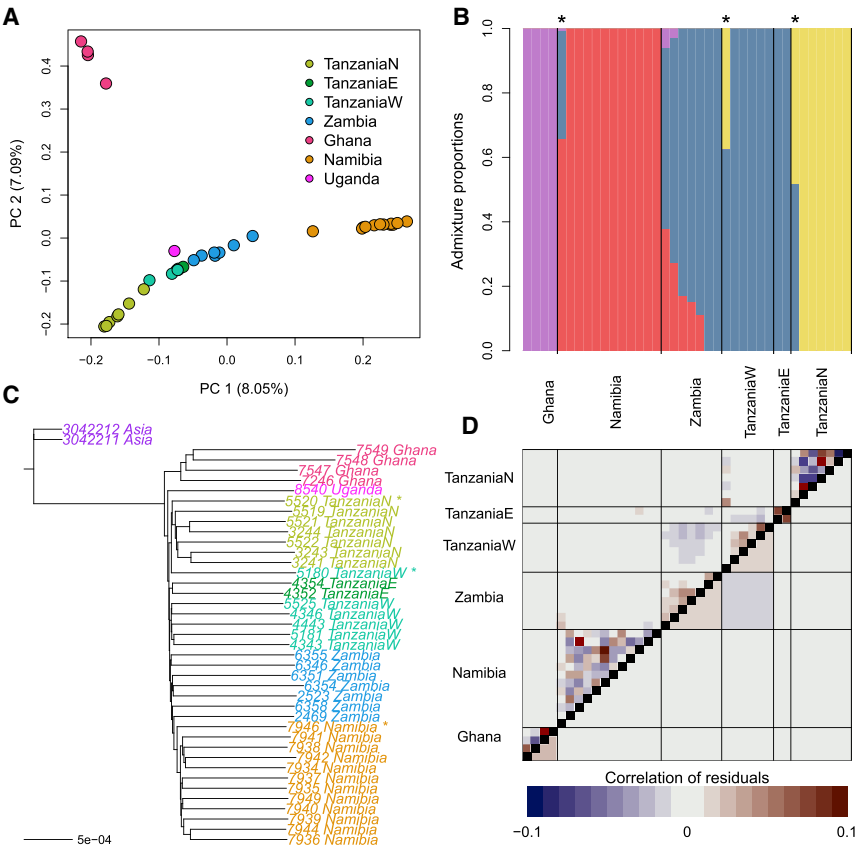


Figure 2. Population structure

(A) PCA of genetic variation in 39 individuals by using PCAngsd,[19] showing PC1 against PC2. Samples are colored by sampling locations. See also Figure S2A for plots showing PC1–PC4.

(B) Ancestry proportions estimated in NGSadmix[20] for K = 4. The single sample from Uganda has been excluded from the analysis. See also Figure S2B for K = 2 and K= 3. The asterisk denotes individuals with distinct ancestry profiles compared with those of the rest of samples from the same population in (B) and (C).

(C) NJ tree based on an identity-by-state (IBS) matrix of 39 individuals calculated in ANGSD[21] and plotted in the R package ape.[22] Scale bar shows the genetic distance.

(D) Evaluation of the admixture proportions shown in (B) inferred by using the evalAdmix program.[23] The upper diagonal shows the correlation of residuals between individuals, and the lower diagonal shows the mean correlation within populations. A positive correlation of residuals is reflective of a bad model fit. Correlation values above and below the color scale are set to dark red and dark blue, respectively. See also Figure S2C for evalAdmix results with K = 2 to K = 4.

size and/or little differentiation between them. However, we note that an evaluation of the admixture model fit using evalAdmix[23] (Figure 2D) indicates that the discrete clustering imposed by NGSadmix is not a good fit to the data, as shown by the presence of numerous non-zero residuals within clusters, consistent with the genetic continuity across geographic space suggested by the PCA (Figure 2A).

To get a better sense of how the samples are genetically clustered, we built a neighbor-joining (NJ) tree based on the 39 individuals from the PCA, as well as two previously published Amur leopards (IDs 3042211 and 3042212).[5] We found that all African leopards form a cluster with a branching pattern consistent with the north-to-south cline observed in the PCA; however, it is evident from the NJ tree that there is very low differentiation between the different sampling locations with short internal branch lengths. Ghana was the most divergent group, whereas the other sampling locations formed a separate cluster with a substructure representing the sampling locations (with the exception of Zambia) in the following order of divergence: Uganda; Tanzania North, East, and West; Zambia; and Namibia (Figure 2C). One sample (ID 5180) stands out as being labeled Tanzania West and clustering with Tanzania North; however, this result is consistent with NGSadmix results, indicating that the sample is a mixture of the Tanzania West and North ancestral sources (Figures 2B and 2C).

Given that the estimated NJ tree largely supports our initial geography-based grouping of the samples, we decided to treat these groupings as populations and quantify the magnitude of genetic differentiation between them by using F_{ST}. However, before doing so, we first confirmed that this population grouping was justifiable by performing a population homogeneity test using *D*-statistics (ABBA-BABA)[25,26] on the low-coverage samples that were not inferred to be admixed in the NGSadmix analysis. Because of low sample sizes, Tanzania East (n = 2) and Uganda (n = 1) were excluded. We tested for relative differences in genetic affinity between all pairs of individuals within a location (H1 and H2) and another sampling location (H3), while using the Amur leopard reference (PanPar1.0)[5] as an outgroup (H4). We found strong signals of substructure in Zambia, where all *D*-statistics were significant (suggesting gene flow with the other sampling locations), in contrast to Ghana, Namibia, Tanzania North, and Tanzania West, where we found zero or one significant test (|Z-score|>3) (Figure 3). This result is consistent with the population structure analyses (Figure 2) and suggests that Ghana, Namibia, Tanzania North, and Tanzania West are meaningful population groupings. We therefore quantified the magnitude of genetic differentiation between populations by estimating pairwise F_{ST}. Overall, we find low levels of differentiation, and the highest values of F_{ST} were 0.118–0.145 (Table 1, lower triangle) between Ghana and the remaining populations, again consistent with the population structure results (Figure 2). In addition, we estimated F_{ST} between all pairs of individuals to factor out any downward bias from potential substructure within the populations. Note that we used the Reich F_{ST} estimator[27] because simulations show that it is robust to low sample size, even down to a single sample from each population (Figure S7B). Because this analysis was based on called genotypes, we restricted it to our five high-coverage samples

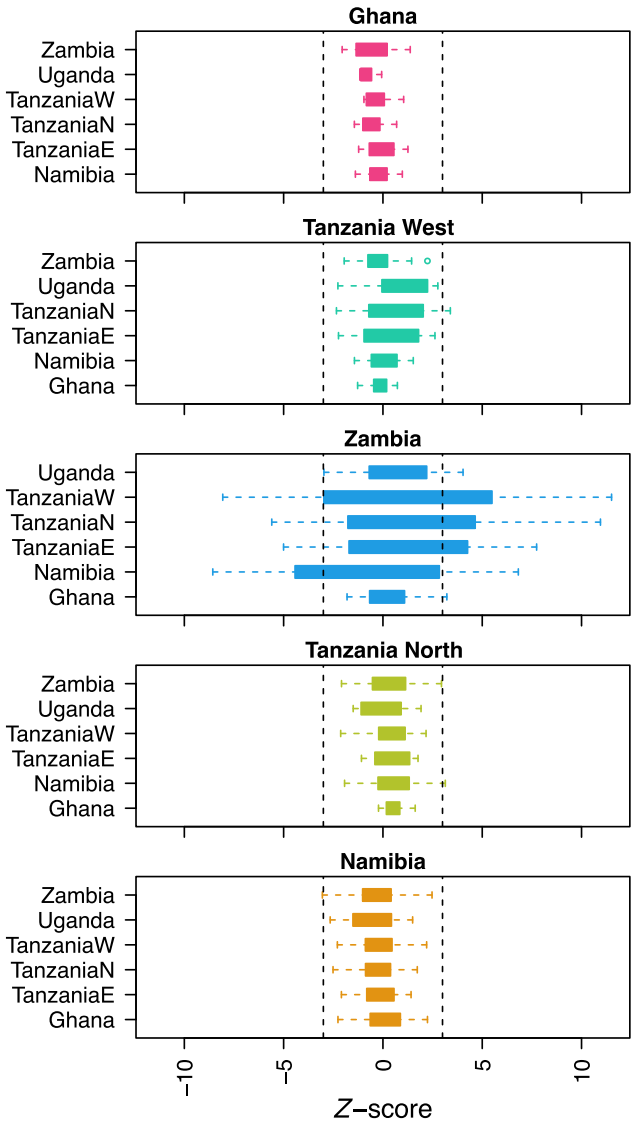


Figure 3. Population homogeneity

D-statistics (ABBA-BABA) testing for homogeneity of designated groups. Each panel is titled by the population being tested (H1 and H2), and the y axis is labeled with the respective remaining African populations (H3). Dashed lines mark the significance thresholds for Z-scores.

(Table 1, upper triangle). The results show a similar range in F_{ST} (0.12–0.17) between Ghana and the remaining populations as the population-based F_{ST} analysis.

Migration within Africa

The migration patterns between the different sampling locations were further investigated by using Estimation of Effective Migration Surfaces (EEMS)[29] (Figure 4). The results fit with an isolation-by-distance model in most parts of the range (Figure S3), suggesting that roughly distance-dependent gene flow is a major determinant of the observed genetic differences. In addition, we find that African leopards show only weak signatures of migration barriers across the latitudinal cline, in particular across the equatorial tropical rainforest. This is in sharp contrast to other African mammals with sub-Saharan distribution, including the lion (*Panthera leo*), for which the dense rainforest is a barrier (Figure 4).

Migration between Africa and Asia

We also tested for signals of post-divergence gene flow between the Amur and African leopards by using *D*-statistics with the domestic cat as an outgroup (H4). We found that all other populations (H2) contrasted against Ghana (H1) show a positive *D*-statistic (Figure 5), suggesting an excess of gene flow in the eastern and southern African lineages with a source represented by, but not necessarily equal to, the Amur leopard (H3). This is consistent with the West African samples being more genetically distant from the Amur leopard than leopards from East and Southern Africa.

Demographic history across the big cats

To investigate the demographic history of the African leopard, we inferred the effective population size through time by using pairwise sequentially Markovian coalescent (PSMC)[30] (Figure 6A). Because PSMC requires called genotypes, we restricted the analysis to the five high-coverage African leopards. We scaled the population size assuming a mutation rate of 10^{-8} per base per year, which has previously been used in studies of big cat species,[10] and a generation time of 7.5 years.[11] For comparison, we applied the same procedure to genomes representing all the other big cats for which we could find publicly available data.[5,10,31] We detected a clear signal of divergence between the African and Amur leopards (Figure 6A), given that their population size trajectories start to differ markedly around 300–400 kya, likely representing the out-of-Africa event. Our PSMC results further show that after the divergence, the African leopards maintained a high effective population size, in contrast to the Amur leopards that went through a bottleneck. Notably, the maintenance of high effective population sizes above 20,000 throughout the Pleistocene is unique for the African leopard in comparison with all other big cats, which all show decreasing effective population sizes in recent times.

Genetic diversity

Finally, we inferred the present-day genetic diversity of African leopards by estimating the genome-wide heterozygosity, i.e., the proportion of heterozygous sites per sample. We inferred on average ~2 heterozygous sites per 1,000 base pairs (bp), with little variation between the various populations of African leopards (Figure 6B). However, we note that the samples from Ghana were within the higher end of the error rate distribution, which could slightly upward-bias the estimates. To investigate the minor differences in heterozygosity between the different populations, we correlated the estimated heterozygosity levels against tracts of runs of homozygosity (ROH) for the five high-coverage samples. We found a negative linear correlation between the total length of ROH and heterozygosity (Figure S4B). This indicates that the majority of variability in genetic diversity among African leopard populations can be explained by differences in recent inbreeding rather than differences in demographic histories.

In line with the distinctive demographic trajectories inferred from PSMC (Figure 6A), the estimated heterozygosity levels for

Table 1. F_{ST} between populations and pairs of individuals

	Ghana	Namibia	TanzaniaN	TanzaniaW	TanzaniaE
Ghana	N/A	0.159	0.169	0.120	0.125
Namibia	0.144	N/A	0.134	0.076	0.080
TanzaniaN	0.145	0.100	N/A	0.069	0.076
TanzaniaW	0.118	0.067	0.051	N/A	0.019
TanzaniaE	N/A	N/A	N/A	N/A	N/A

Lower triangle shows the estimated F_{ST} between pairs of populations. The upper triangle shows the estimated F_{ST} between pairs of high-coverage individuals. Tanzania East (TanzaniaE) was excluded from the population-based analysis. Figure S7B shows that the F_{ST} between pairs of individuals is not biased. The population based F_{ST} was based on 2d-SFS from all individuals from each population estimated with realSFS,[28] whereas the individual F_{ST} was based on the 2d-SFS from called genotypes from a high-coverage individual from each population.

the African leopards are 3-fold higher than those in the Amur leopards. They are also markedly higher than for any other big cat species, and most felid species in general (Figure 6A; Table S6), with the exception of two distantly related felids: the leopard cat (*Prionailurus bengalensis*) and to a lesser degree the puma (*Puma concolor*).

DISCUSSION

Our results provide new insights into the evolutionary history and molecular ecology of the leopard in its biogeographic cradle, the African continent. We found that the genetic diversity of the leopard was previously severely underestimated due to being characterized in Amur leopards, which have an unrepresentative population history for the species. We show that, in addition to suffering recent population declines,[32] Amur leopards also carry a genetic legacy of the out-of-Africa event occurring in the middle Pleistocene (Figure 6A) and possibly a number of subsequent founder events. Lower genetic diversity is also frequently observed at the edge of species ranges because of somewhat extreme demographic histories.[33]

In contrast to the Amur leopards, African leopards retained large effective population sizes after the split. Furthermore, African leopards consistently maintained much higher population sizes than all other big cats throughout the Pleistocene (Figure 6A). Declining population sizes were also recently inferred in the puma[34] and the cheetah (*Acinonyx jubatus*).[35] Consistent with these differences in effective population size histories, the African leopards have by far the highest genetic diversity not only among big cats (Figure 6B) but among wild cats in general, matched only by the leopard cat, a small felid mesopredator species with a wide distribution in southern and eastern Asia.[34,36]

Interestingly, our results suggest that after the divergence, the African populations remained genetically connected with the Asian populations, given that we detected gene flow with a population represented by the Amur leopard lineage and the African leopards (Figure 5). The extent of Asian ancestry is not equal between African populations, suggesting either ancestral population structure before the split or a backflow into the ancestral branch of the eastern and southern African populations. Uphyrkina et al.[14] found that the South Arabian leopard (*P. p. nimr*) could not be consistently placed with either African or Asian

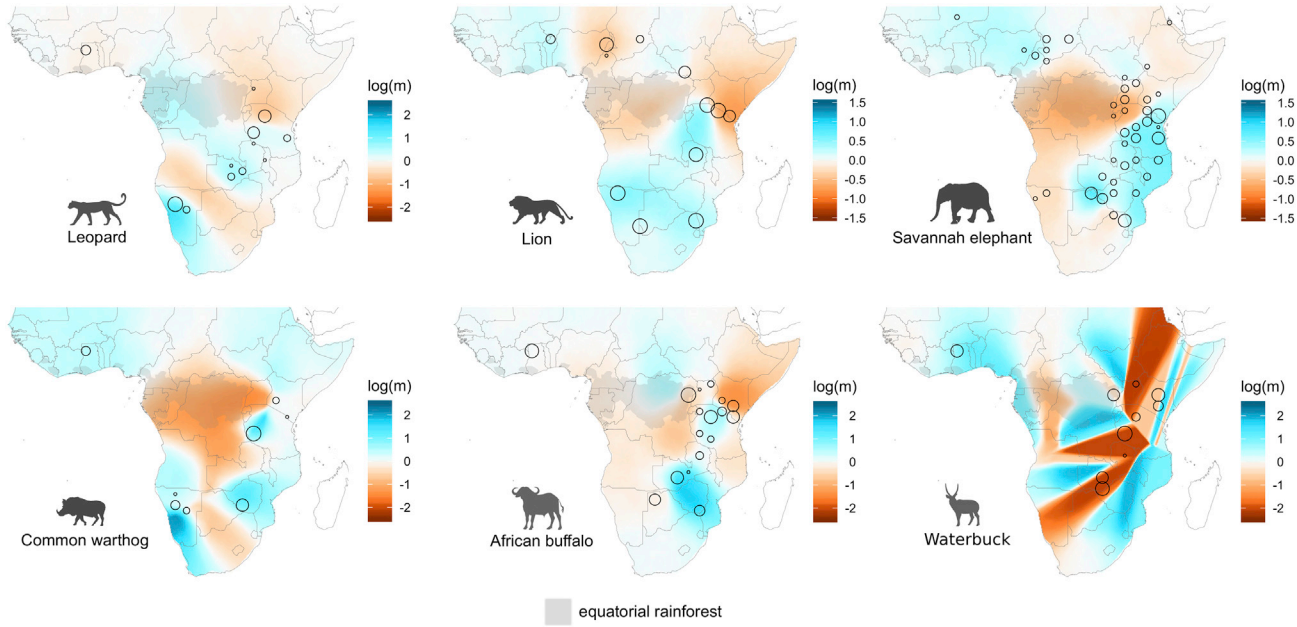


Figure 4. Reconstruction of migration surfaces

Estimation of effective migration surfaces (EEMS)[29] was used to identify barriers to gene flow and regions of sustained genetic connectivity. The color scale of blue to dark orange represents the log10-transformed effective migration rates (m); the blue indicates areas with relatively higher migration, and orange indicates areas with relatively lower migration. Circles represent the demes (subpopulations) rather than the predefined sampling locations. The lion and elephant plots have smaller scales to make the coloring clearer because of a lower resolution in the microsatellite data. The equatorial rainforest is highlighted with gray shading (according to Mayaux et al.[18]) to show that, unlike in the other species, it does not constitute a barrier for gene flow in leopards. See also Figure S3.

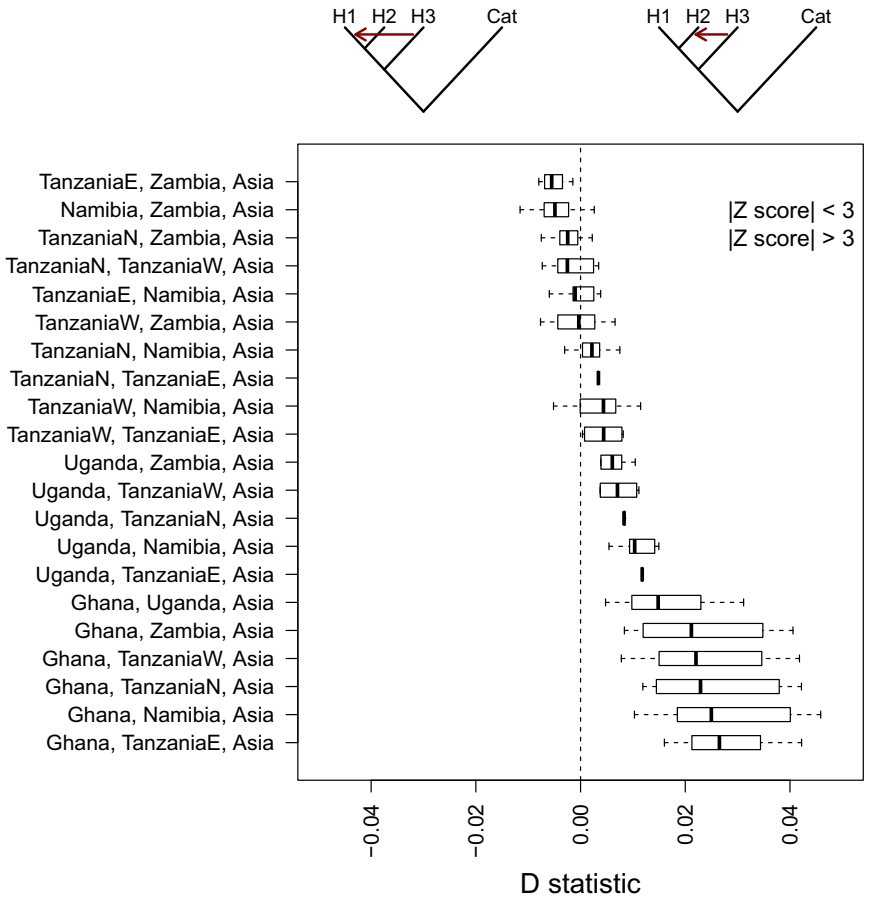


Figure 5. *D*-statistics with low coverage samples and Amur leopard

All pairs of individuals assigned to each possible pair of African locations are grouped as H1 and H2 with the Asian leopard sample with ID 3042211 as H3. The domestic cat was used as an outgroup (H4). Comparisons with a significant *D*-statistic (absolute Z score above 3) are plotted as golden points.

leopards, which would be in accordance with genetic connectivity between a proto-eastern African leopard population and those of the Arabian Peninsula and the Middle East. Our results therefore suggest Pleistocene-era intercontinental genetic connectivity comparable to the level observed in another versatile carnivore, the spotted hyena.[37]

The topology of our genome-wide NJ tree is consistent with a southward expansion of the eastern African lineage after divergence with Ghana (Figure 2C); however, the internal branch lengths are very short, suggesting that any such pattern is only tentative and obscured by pervasive gene flow. Overall, we found signatures of an isolation-by-distance model in most parts of the range, but this pattern is attenuated by extensive gene flow on a continent-wide scale within Africa (Figure S3). Concordantly, we found that African leopards show at best weak signatures of dispersal barriers across the latitudinal cline (Figure 4). Although the character of sampling and low density of samples from western and northern Africa prevent us from resolving genetic connectivity patterns on a finer scale, the continent-level pattern in leopards clearly stands out compared with that of other co-distributed African mammals by presence of gene flow across the equatorial rainforest. The equatorial tropical rainforest is one of the main drivers of diversification for many African mammals,[38] but we found no evidence of it being a barrier for leopards, underlining their exceptional habitat tolerance. Accordingly, the leopard also stands out as having low levels of population structure compared with that of the majority of other species in the African large-mammal savanna guild that show strong regional structuring.[38–41] The average pairwise F_{ST} between populations was 0.10, which is considerably lower than corresponding values for other co-distributed African carnivores (cheetah: 0.29[42] and lion: 0.25[43]). Although genome-wide data can be expected to capture finer-scale structure, our results are consistent with early studies based on microsatellite data and partial mitochondrial sequences, which reported limited genetic structure within the African continent.[14,15] However, a recent and densely sampled study contrastingly found strong structure in Africa.[17] This discrepancy is possibly explained by the use of mitochondrial data,[17] which only reflects the female lineage and fails to capture the male-biased dispersal in leopards.[44]

As typical apex predators, the big cats of the genus *Panthera* are specialized hypercarnivores depending on large prey species.[45,46] Being at the top of the trophic pyramid,[3,4] carnivores tend to have lower effective population sizes, and are thus more affected by environmental fluctuations and habitat fragmentation in periods when the climatic conditions are not favorable.[47,48] As a consequence, carnivores are prone to increased genetic drift and low genetic diversity compared with herbivores,[5,6,49] leading to an increased risk of extinction,[4,50] with Holliday and Steppan[45] labeling highly specialized carnivory as an evolutionary "dead end." In stark contrast to its congeners, we find that the African leopard is an exceptionally adaptable apex predator. High mobility,[51,52] habitat versatility,[8,9] and dietary generalism[13,53] have buffered the long-term high effective population sizes in the African leopards by making them less sensitive to habitat fragmentation and environmental fluctuations during the Pleistocene climatic cycles. In this light, our results are surprising, but they are in line with the biology of the leopard being defined by a broader dietary and habitat niche than any other big cat. We argue that the African leopard might constitute an evolutionary anomaly with a better chance of long-term survival than other *Panthera* species. Future studies involving more extensive sampling throughout the leopard range will resolve how current genetic diversity is connected with

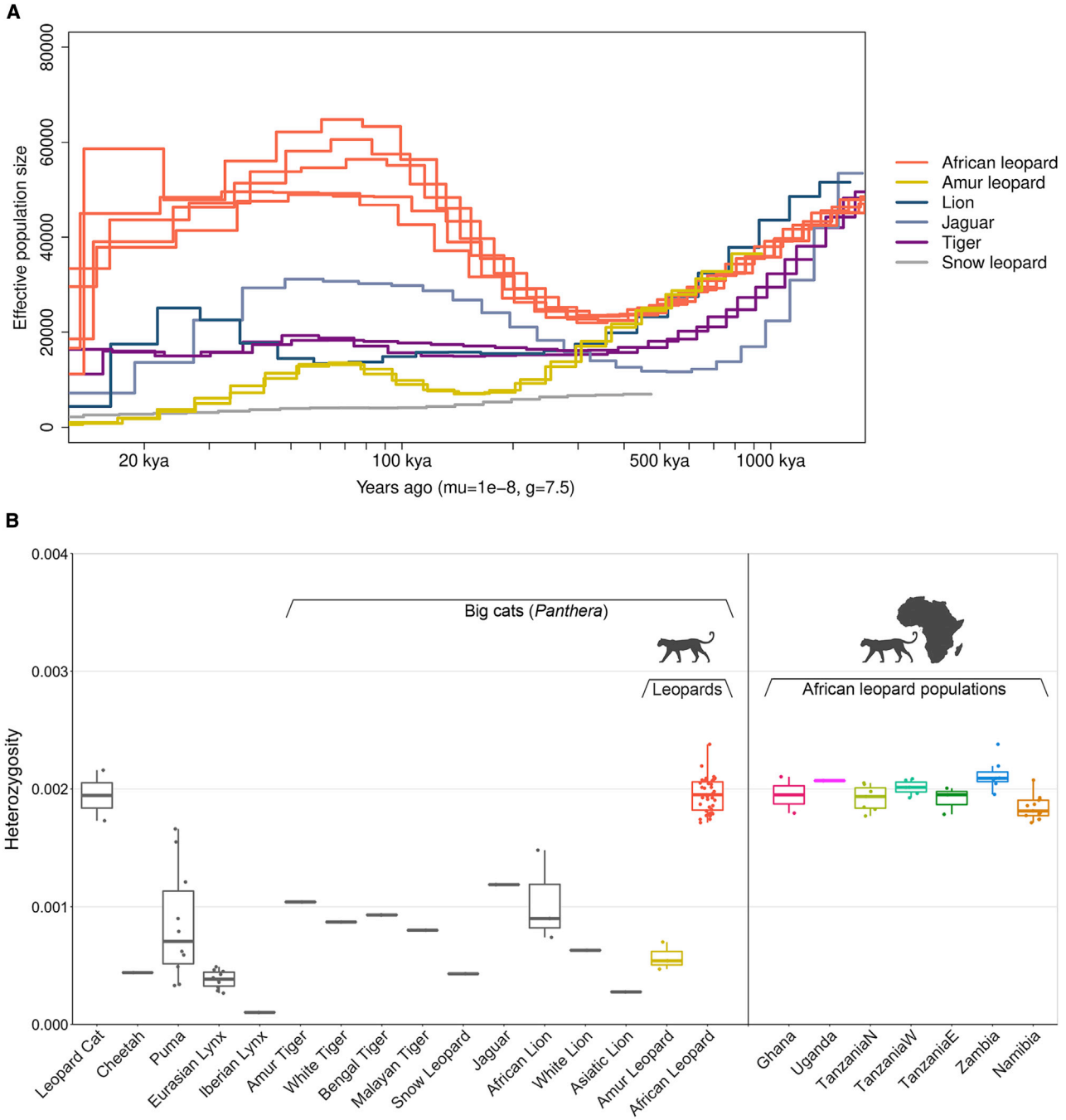


Figure 6. Demographic history and genetic diversity

(A) Effective population sizes were reconstructed by using pairwise sequentially Markovian coalescent (PSMC[30]) assuming a mutation rate, mu, of 10^{-8} and a generation time, g, of 7.5 years. African leopards have a distinct demographic history compared with that of the Amur leopard and the other big cats. See also Figure S7A for PSMC with a generation time of 5 years.

(B) Estimates of genome-wide heterozygosity inferred in ANGSD[21] are compared between the African leopard populations, and also between all felid species with available whole-genome estimates. See also Figures S4, S6, S7A and Table S6.

demographic history. We speculate that the Amur leopard will most likely not be representative of other Asian leopard subspecies because of its recent bottleneck[5] and its position at the extremity of the leopard geographic range (Figure S1).

Despite their resilience, leopards have lost 48%–67% of their original distribution in Africa within the last ~300 years[13] (Figure S1), and their range is contracting at a similar rate to some of the other big cats.[54,55] Even though our results highlight that the

leopards might be among the most adaptable of apex predators, and therefore relatively robust to environmental change and disturbance, anthropogenic activities can pose an unprecedented threat to their survival. Until recently, these have not had a major effect on leopard populations in Africa, but their potential effect is evident in the other leopard subspecies, particularly the critically endangered Amur leopard, which has been strongly affected by the lack of available prey, illegal hunting, and fragmentation of habitat.[11] In addition, unlike species that went through periods of low population size, African leopards have had constantly high population sizes and have not endured bottlenecks, which would have purged strongly deleterious variation from the gene pool. African leopards might therefore harbor a larger number of strongly deleterious mutations at low population frequencies. These can increase in frequency as a result of population contractions, placing the African leopard at risk of inbreeding depression.[56–58]

In conclusion, we correct the existing bias in the assessment of leopard genetic diversity and show that African leopards have exceptionally high genetic diversity for their trophic position, coupled with high continent-wide genetic connectivity in one of their biogeographic strongholds. Also, on the basis of our observations of low population differentiation and lack of migration barriers across Africa, we hypothesize that these unusual genetic features are caused by the resilience, adaptability, and high mobility of the leopard. This study represents a rare positive narrative in conservation genetics and, to our knowledge, one of the most convincing couplings of a species' generalist ecology with its correspondingly exceptional long-term evolutionary dynamics and genetic features.

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 SUBJECT DETAILS
- METHOD DETAILS
  - DNA extraction
  - Sequencing
  - Mapping
- QUANTIFICATION AND STATISTICAL ANALYSES
  - Reference Quality filtering: Mappability and RepeatMasker
  - Reference Quality filtering: Global Sequencing Depth filter
  - Reference Quality filtering: Autosomal and sex-linked scaffold identification
  - Reference Quality filtering: Excessive heterozygosity filter
  - Reference Quality filtering: Summarizing reference quality filters
  - Sample quality filtering: Error rates estimation
  - Sample quality filtering: BLAST analyses of mtDNA
  - Sample quality filtering: Identification of related and duplicated samples
  - Genotype likelihoods, SNPs, and genotype calling
  - PCA analyses
  - NGSadmix
  - Runs of homozygosity (ROH) analysis
  - *D*-statistics
  - Site frequency spectrum
  - SFS-based analyses: F_{ST} estimation
  - SFS-based analyses: Estimation of heterozygosity
  - PSMC
  - Identity-by-state (IBS)
  - IBS-based analyses: Neighbor-joining (NJ) tree
  - IBS-based analyses: Estimation of Effective Migration Surfaces (EEMS)

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2021.01.064.

ACKNOWLEDGMENTS

The authors would like to thank Peter Arctander for providing access to his existing collection of leopard samples that were analyzed in this study. We also thank David Moyer for contributing samples to the collection. The wildlife management authorities of Ghana, Namibia, Tanzania, Uganda, and Zambia are thanked for collaborating with Peter Arctander in the sample collection. P.F. was supported by the Innovation Fund Denmark, Candys Foundation, and the Alfred Benzon Fund. R.H. was supported by a Villum Foundation Young investigator grant (VKR023447) and a Danish Research Council Sapere Aude grant (DFF8049-00098B). G.G.E., J.M., A.A., and K.H. were supported by a Lundbeck Foundation grant (R215-2015-4174). I.M. and X.L. were supported by a Villum Foundation Young Investigator grant (19114), R.R.d.F. thanks the Danish National Research Foundation for its funding of the Center for Macroecology, Evolution, and Climate (DNRF96).

AUTHOR CONTRIBUTIONS

H.R.S., A.A., R.H., and I.M. conceived and designed the study. P.P., G.G.E., X.L., C.N., R.K.W., C.G.S., L.Q., P.F., J.M., F.F.S., A.B.-O., C.H.F.J., R.R.d.F., H.R.S., A.A., R.H., I.M., and K.H. analyzed the data. P.P., A.A., R.H., I.M., and K.H. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 8, 2020
Revised: November 13, 2020
Accepted: January 19, 2021
Published: February 25, 2021

REFERENCES

1. Voigt, W., Perner, J., Davis, A.J., Eggers, T., Schumacher, J., Bährmann, R., Fabian, B., Heinrich, W., Köhler, G., Lichter, D., et al. (2003). Trophic levels are differentially sensitive to climate. Ecology *84*, 2444–2453.
2. Cheng, B.S., Komoroske, L.M., and Grosholz, E.D. (2017). Trophic sensitivity of invasive predator and native prey interactions: integrating environmental context and climate change. Funct. Ecol. *31*, 642–652.
3. Stier, A.C., Samhouri, J.F., Novak, M., Marshall, K.N., Ward, E.J., Holt, R.D., and Levin, P.S. (2016). Ecosystem context and historical contingency in apex predator recoveries. Sci. Adv. *2*, e1501769.

4. Purvis, A., Gittleman, J.L., Cowlishaw, G., and Mace, G.M. (2000). Predicting extinction risk in declining species. Proc. Biol. Sci. *267*, 1947–1952.
5. Kim, S., Cho, Y.S., Kim, H.-M., Chung, O., Kim, H., Jho, S., Seomun, H., Kim, J., Bang, W.Y., Kim, C., et al. (2016). Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. Genome Biol. *17*, 211.
6. Díez-Del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M.T.P., and Dalén, L. (2018). Quantifying Temporal Genomic Erosion in Endangered Species. Trends Ecol. Evol. *33*, 176–185.
7. Nowak, R.M., and Walker, E.P. (1999). Walker's Mammals of the World (JHU Press).
8. Nowell, K., and Jackson, P. (1996). Wild cats: status survey and conservation action plan (Switzerland: IUCN Gland).
9. Hunter, L., Henschel, P., and Ray, J.C. (2013). *Panthera pardus*. In Mammals of Africa Volume V: Carnivores, Pangolins, Equids and Rhinoceroses, J. Kingdon, D. Happold, T. Butynski, M. Hoffmann, M. Happold, and J. Kalina, eds. (Bloomsbury Publishing PLC London, UK), pp. 159–168.
10. 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.
11. Stein, A.B., Athreya, V., Gerngross, P., Balme, G., Henschel, P., Karanth, U., et al. (2020). *Panthera pardus* (amended version of 2019 assessment). The IUCN Red List of Threatened Species 2020, e.T15954A163991139.
12. Paijmans, J.L.A., Barlow, A., Förster, D.W., Henneberger, K., Meyer, M., Nickel, B., Nagel, D., Worsøe Havmøller, R., Baryshnikov, G.F., Joger, U., et al. (2018). Historical biogeography of the leopard (*Panthera pardus*) and its extinct Eurasian populations. BMC Evol. Biol. *18*, 156.
13. Jacobson, A.P., Gerngross, P., Lemeris, J.R., Jr., Schoonover, R.F., Anco, C., Breitenmoser-Würsten, C., Durant, S.M., Farhadinia, M.S., Henschel, P., Kamler, J.F., et al. (2016). Leopard (*Panthera pardus*) status, distribution, and the research efforts across its range. PeerJ *4*, e1974.
14. Uphyrkina, O., Johnson, W.E., Quigley, H., Miquelle, D., Marker, L., Bush, M., and O'Brien, S.J. (2001). Phylogenetics, genome diversity and origin of modern leopard, *Panthera pardus*. Mol. Ecol. *10*, 2617–2633.
15. Miththapala, S., Seidensticker, J., and O'Brien, S.J. (1996). Phylogeographic Subspecies Recognition in Leopards (*Panthera pardus*). Molecular Genetic Variation. Conserv. Biol. *10*, 1115–1132.
16. Kitchener, A.C., Breitenmoser-Würsten, C., Eizirik, E., Gentry, A., Werdelin, L., Wilting, A., Yamaguchi, N., Abramov, A.V., Christiansen, P., Driscoll, C., et al. (2017). A revised taxonomy of the Felidae: The final report of the Cat Classification Task Force of the IUCN Cat Specialist Group (Cat News).
17. Anco, C., Kolokotronis, S.-O., Henschel, P., Cunningham, S.W., Amato, G., and Hekkala, E. (2018). Historical mitochondrial diversity in African leopards (*Panthera pardus*) revealed by archival museum specimens. Mitochondrial DNA A. DNA Mapp. Seq. Anal. *29*, 455–473.
18. Mayaux, P., Pekel, J.-F., Desclée, B., Donnay, F., Lupi, A., Achard, F., Clerici, M., Bodart, C., Brink, A., Nasi, R., and Belward, A. (2013). State and evolution of the African rainforests between 1990 and 2010. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120300.
19. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.
20. Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702.
21. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.
22. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.
23. Garcia-Erill, G., and Albrechtsen, A. (2020). Evaluation of model fit of inferred admixture proportions. Mol. Ecol. Resour. *20*, 936–949.
24. Puechmaille, S.J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol. Ecol. Resour. *16*, 608–627.
25. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.
26. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252.
27. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.
28. 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.
29. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100.
30. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
31. Cho, Y.S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., Oh, S., Kim, H.-M., Jho, S., Kim, S., et al. (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. Nat. Commun. *4*, 2433.
32. Uphyrkina, O., Miquelle, D., Quigley, H., Driscoll, C., and O'Brien, S.J. (2002). Conservation genetics of the Far Eastern leopard (*Panthera pardus orientalis*). J. Hered. *93*, 303–311.
33. Eckert, C.G., Samis, K.E., and Lougheed, S.C. (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol. Ecol. *17*, 1170–1188.
34. Saremi, N.F., Supple, M.A., Byrne, A., Cahill, J.A., Coutinho, L.L., Dalén, L., Figueiró, H.V., Johnson, W.E., Milne, H.J., O'Brien, S.J., et al. (2019). Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. Nat. Commun. *10*, 4769.
35. Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A.A., Krasheninnikova, K., Kliver, S., Schmidt-Küntzel, A., Koepfli, K.-P., Johnson, W., et al. (2015). Genomic legacy of the African cheetah, *Acinonyx jubatus*. Genome Biol. *16*, 277.
36. Corbet, G.B., Hill, J.E., and Hill, J.E. (1992). The Mammals of the Indomalayan Region: A Systematic Review (Oxford University Press on Demand).
37. Westbury, M.V., Hartmann, S., Barlow, A., Preick, M., Ridush, B., Nagel, D., Rathgeber, T., Ziegler, R., Baryshnikov, G., Sheng, G., et al. (2020). Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. Sci Adv *6*, eaay0456.
38. Lorenzen, E.D., Heller, R., and Siegismund, H.R. (2012). Comparative phylogeography of African savannah ungulates. Mol. Ecol. *21*, 3656–3670.
39. de Manuel, M., Barnett, R., Sandoval-Velasco, M., Yamaguchi, N., Garrett Vieira, F., Zepeda Mendoza, M.L., Liu, S., Martin, M.D., Sinding, M.S., Mak, S.S.T., et al. (2020). The evolutionary history of extinct and living lions. Proc. Natl. Acad. Sci. USA *117*, 10927–10934.
40. Nyakaana, S., Arctander, P., and Siegismund, H.R. (2002). Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. Heredity *89*, 90–98.
41. 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.
42. Prost, S., Machado, A.P., Zumbroich, J., Preier, L., Mahtani-Williams, S., Guschanski, K., Brealey, J.C., Fernandes, C., Vercammen, P., Godsall-Bottriell, L., et al. (2020). Conservation genomic analyses of African and Asiatic cheetahs (Acinonyx jubatus) across their current and historic species range. bioRxiv, 2020.02.14.949081.
43. Smitz, N., Jouvenet, O., Ambwene Ligate, F., Crosmary, W.-G., Ikanda, D., Chardonnet, P., Fusari, A., Meganck, K., Gillet, F., Melletti, M., and Michaux, J.R. (2018). A genome-wide data assessment of the African

lion (*Panthera leo*) population genetic structure and diversity in Tanzania. PLoS ONE *13*, e0205395.

44. Fattebert, J., Balme, G., Dickerson, T., Slotow, R., and Hunter, L. (2015). Density-dependent natal dispersal patterns in a leopard population recovering from over-harvest. PLoS ONE *10*, e0122355.
45. Holliday, J.A., and Steppan, S.J. (2004). Evolution of hypercarnivory: the effect of specialization on morphological and taxonomic diversity. Paleobiology *30*, 108–128.
46. Shrestha, B., Reed, J.M., Starks, P.T., Kaufman, G.E., Goldstone, J.V., Roelke, M.E., O'Brien, S.J., Koepfli, K.-P., Frank, L.G., and Court, M.H. (2011). Evolution of a major drug metabolizing enzyme defect in the domestic cat and other felidae: phylogenetic timing and the role of hypercarnivory. PLoS ONE *6*, e18046.
47. Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: a review. Genet. Res. *66*, 95–107.
48. Frankham, R., Ballou, J.D., and Briscoe, D.A. (2010). Introduction to Conservation Genetics, 2nd ed. (Cambridge: Cambridge University Press). https://doi.org/10.1017/CBO9780511809002.
49. Hansen, C.C.R., Hvilsom, C., Schmidt, N.M., Aastrup, P., Van Coeverden de Groot, P.J., Siegismund, H.R., et al. (2018). The muskox lost a substantial part of Its genetic diversity on its long road to Greenland. Curr. Biol. *28*, 4022–4028.
50. Van Valkenburgh, B., Wang, X., and Damuth, J. (2004). Cope's rule, hypercarnivory, and extinction in North American canids. Science *306*, 101–104.
51. Fattebert, J., Dickerson, T., Balme, G., Slotow, R., and Hunter, L. (2013). Long-distance natal dispersal in leopard reveals potential for a three-country metapopulation. Afr. J. Wildl. Res. *43*, 61–67.
52. Ray-Brambach, R.R., Stommel, C., and Rödder, D. (2018). Home ranges, activity patterns and habitat preferences of leopards in Luambe National Park and adjacent Game Management Area in the Luangwa Valley, Zambia. Mamm. Biol. *92*, 102–110.
53. Hayward, M.W., Henschel, P., O'brien, J., Hofmeyr, M., Balme, G., and Kerley, G.I.H. (2006). Prey preferences of the leopard (*Panthera pardus*). J. Zool. (Lond.) *270*, 298–313.
54. Mahmood, T., Younas, A., Akrim, F., Andleeb, S., Hamid, A., and Nadeem, M.S. (2019). Range contraction of snow leopard (*Panthera uncia*). PLoS ONE *14*, e0218460.
55. Quigley, H., Foster, R., Petracca, L., Payan, E., Salom, R., and Harmsen, B. (2017). *Panthera onca* (errata version published in 2018). The IUCN Red List of Threatened Species 2017, e.T15953A123791436.
56. Robinson, J.A., Räikkönen, J., Vucetich, L.M., Vucetich, J.A., Peterson, R.O., Lohmueller, K.E., and Wayne, R.K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv *5*, eaau0757.
57. Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J.L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J.Y., Prieto, P., Quesada, V., et al. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biol. *17*, 251.
58. Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.
59. Gaspar, J.M. (2018). NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics *19*, 536.
60. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
61. 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.
62. Pockrandt, C., Alzamel, M., Iliopoulos, C.S., and Reinert, K. (2020). GenMap: ultra-fast computation of genome mappability. Bioinformatics *36*, 3687–3692.
63. Smit, A.F.A. (2004). Repeat-Masker Open-3.0. http://www.repeatmasker.org. Available at: https://ci.nii.ac.jp/naid/10029514778/ [Accessed August 17, 2020].
64. Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., and Schäffer, A.A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics *24*, 1757–1764.
65. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
66. 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.
67. 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.
68. Köster, J., and Rahmann, S. (2012). Snakemake–a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522.
69. Pontius, J.U., Mullikin, J.C., Smith, D.R., Lindblad-Toh, K., Gnerre, S., Clamp, M., Chang, J., Stephens, R., Neelam, B., Volfovsky, N., et al. (2007). Agencourt Sequencing Team; NISC Comparative Sequencing Program Initial sequence and comparative analysis of the cat genome. Genome Res. *17*, 1675–1689.
70. Wei, L., Wu, X., Zhu, L., and Jiang, Z. (2011). Mitogenomic analysis of the genus *Panthera*. Sci. China Life Sci. *54*, 917–930.
71. Meisner, J., and Albrechtsen, A. (2019). Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. Mol. Ecol. Resour. *19*, 1144–1152.
72. 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.
73. Waples, R.K., Albrechtsen, A., and Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Mol. Ecol. *28*, 35–48.
74. 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.
75. Busing, F.M.T.A., Meijer, E., and Leeden, R.V.D. (1999). Delete-m jackknife for unequal m. Stat. Comput. *9*, 3–8.
76. Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., and Weir, B.S. (2015). CONSERVATION. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science *349*, 84–87.
77. Bertola, L.D., Tensen, L., van Hooft, P., White, P.A., Driscoll, C.A., Henschel, P., Caragiulo, A., Dias-Freedman, I., Sogbohossou, E.A., Tumenta, P.N., et al. (2015). Autosomal and mtDNA Markers Affirm the Distinctiveness of Lions in West and Central Africa. PLoS ONE *10*, e0137975.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
RNase A	QIAGEN	Cat# 19101
Critical Commercial Assays		
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69506
AxyPrep MAG PCR Clean-Up Kit	Axygen	MAG-PCR-CL-50
QIAquick Gel Extraction kit	QIAGEN	Cat# 28706X4
Deposited Data		
Raw sequencing reads	This study	ENA accession number: PRJEB41230
Software and Algorithms		
NGmerge	Gaspar[59]	https://github.com/jsh58/NGmerge
BWA-mem v0.7.17	Li and Durbin[60]	http://bio-bwa.sourceforge.net/
Samtools v1.9	Li et al.[61]	http://www.htslib.org/
GENMAP v1.2.0	Pockrandt et al.[62]	https://github.com/cpockrandt/genmap
RepeatMasker v4.0.8	Smit[63]	http://www.repeatmasker.org/
ANGSD	Korneliussen[21]	http://www.popgen.dk/angsd/index.php/ANGSD
PCAngsd v0.973	Meisner and Albrechtsen[19]	http://www.popgen.dk/software/index.php/PCAngsd
megaBLAST	Morgulis et al.[64]	https://blast.ncbi.nlm.nih.gov
GATK	McKenna et al.[65]	https://gatk.broadinstitute.org/hc/en-us
NGSAdmix	Skotte et al.[20]	http://www.popgen.dk/software/index.php/NgsAdmix
EvalAdmix	Garcia-Erill and Albrechtsen[23]	https://github.com/GenisGE/evalAdmix
PLINK v1.9	Purcell et al.[66]	https://www.cog-genomics.org/plink/
realSFS v0.931	Nielsen et al.[28]	http://www.popgen.dk/angsd/index.php/RealSFS
Bcftools v1.9	Li[67]	http://samtools.github.io/bcftools/
PSMC	Li and Durbin[30]	https://github.com/lh3/psmc
R package ape	Purcell et al.[66]	https://cran.r-project.org/package=ape
EEMS	Petkova et al.[29]	https://github.com/dipetkov/eems
RStudio v1.2.5033	RStudio Team	https://rstudio.com/
make_eems_plots script	Petkova et al.[29]	https://github.com/dipetkov/reemsplots2
Snakemake	Köster and Rahmann[68]	https://snakemake.readthedocs.io/en/stable/
Other		
Panthera pardus reference genome PanPar1.0	Kim et al.[5]	GenBank: GCF_001857705.1
Domestic cat reference genome Felis_catus_9.0	Pontius et al.[69]	GenBank: GCA_000181335.4
Panthera pardus mitogenome	Wei et al.[70]	GenBank: NC_010641.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kristian Hanghøj (kristian.hanghoej@gmail.com).

Materials availability

This study did not generate new unique reagents

Data and code availability

The accession number for the sequencing data reported in this paper is ENA: PRJEB41230. The code used for analyses in this project is available on the project's github: https://github.com/KHanghoj/leopardpaper. Some pipelines were developed using snakemake.[68]

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We analyzed fifty-three tissue samples of African leopards kindly provided from an existing collection of Peter Arctander. These samples were collected between 1993–1998 in Ghana, Namibia, Tanzania, Uganda, and Zambia (Table S1), with the assistance of the local wildlife management authorities and in compliance with the local and international legislation. Most samples were from skin tissue and were kept in a DMSO buffer in the field and stored at −20°C as soon as possible. The samples were subsequently transferred to a −80°C freezer for long-term storage.

METHOD DETAILS

DNA extraction

For DNA extraction, we used the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA, USA), following the manufacturer's protocol. 10uL RNase A was added to get RNA-free genomic DNA. DNA concentrations were measured with a Qubit 2.0 Fluorometer and with a Nanodrop. We subsequently used gel electrophoresis to ascertain genomic DNA quality.

Sequencing

After DNA extraction, 1 μg genomic DNA was randomly fragmented by Covaris (350 bp on average), followed by purification by AxyPrep Mag PCR clean up kit. The fragments were end repaired by End Repair Mix and purified afterward. The repaired DNAs were combined with A-Tailing Mix, then the Illumina adaptors were ligated to the DNA adenylate 3′ ends, followed by product purification. Size selection was performed targeting insert size of 350 base pairs. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the Adaptor-ligated DNA fragments. After purification, the libraries were assessed by the Agilent Technologies 2100 Bioanalyzer and ABI StepOnePlus Realtime PCR System.

All samples were sequenced in paired-end 2x150 bp mode, 47 of these to approximately 2-5X depth of coverage on the Illumina NovaSeq platform, and six samples were sequenced to approximately 15-20X depth of coverage on the Illumina HiSeq2500 platform. In total, 7.2 billion raw reads were generated and analyzed (Table S2).

Mapping

Cleaned paired-end reads were collapsed when read termini overlapped by at least 11 base pairs (bp) using NGmerge[59] (options: -b 11 -g -f). Collapsed reads and uncollapsed paired-end reads were mapped separately with BWA-MEM (v0.7.17; default settings)[60] to: 1) a *Panthera pardus* reference genome (PanPar1.0; GenBank: GCF_001857705.1,[5] representing a female zoo specimen of an Amur leopard), and 2) a reference genome from a female domestic cat (Felis_catus_9.0; GenBank: GCA_000181335.4).[69] Read duplicates (markdup, default settings) were marked and removed (-F 3852) using Samtools (v1.9).[61] We retained mapped collapsed reads, and properly paired and mapped paired-end reads. All reads with mapping quality below 30 were excluded from all downstream analysis. For mapping statistics, see Table S2.

QUANTIFICATION AND STATISTICAL ANALYSES

Reference Quality filtering: Mappability and RepeatMasker

We estimated the mappability of each site in the reference genomes using GENMAP (v1.2.0).[62] Mappability scores were computed with 100 bp k-mers, allowing two mismatches (-K 100 -E 2), with remaining settings set to default. We excluded all sites with a mappability score less than one. We also excluded low complexity and repeat sequences as identified with RepeatMasker (v.4.0.8, sensitive mode, -engine wublast -s -no_is -cutoff 255 -frag 20000, http://www.repeatmasker.org/).[63] Lastly, we removed all scaffolds smaller than 1 MB.

Reference Quality filtering: Global Sequencing Depth filter

We calculated the global depth (read count across all samples) for each site using ANGSD.[21] Next, we calculated the lower and upper 1% percentile global depth threshold and excluded all sites outside this range. This analysis was done for all samples as well as separately for the low-coverage and high-coverage samples (Figure S5A).

Reference Quality filtering: Autosomal and sex-linked scaffold identification

To identify sex-linked scaffolds, we first calculated the average depth for each scaffold for each sample normalized by the average depth of the five largest scaffolds. Based on the normalized depths, we then performed a principal component analysis (PCA) and identified two main clusters of samples, likely representing males and females (Figure S5B). Finally, considering that in leopards females have two X chromosomes and males have a single X chromosome, we designated and excluded putative sex-linked scaffolds where the mean difference was greater than 0.4 between the two clusters. In addition, we excluded all scaffolds with an average normalized depth greater than 1.1 or lower than 0.9 across all samples (Figure S5C). This resulted in 23 scaffolds being identified as X chromosome-linked (Figure S5D) and excluded from further analyses.

Reference Quality filtering: Excessive heterozygosity filter

Several factors related to genomic repeats, like the presence of copy number variation between the reference genome and the analyzed samples, can lead to mis-mapping of reads originating from multiple genome locations to a single location. Any differences in these locations will result in inferred sites with an excess of heterozygosity that can be used to identify and mask these regions. We computed a preliminary genotype likelihoods file for common polymorphic sites (MAF $\geq$ 0.05 and SNP p value < 10^{-6}) restricted to data with a base quality of at least 30.[21] Using these genotype likelihoods as input to PCAngsd,[19] we calculated per-site inbreeding coefficients (F), ranging from -1 where all samples are heterozygous to 1 where all samples are homozygous, and performed a Hardy-Weinberg equilibrium (HWE) likelihood ratio test accounting for population structure.[71] To obtain individual allele frequencies in PCAngsd, three principal components that account for population structure were considered. Based on the per-site inbreeding coefficients, windows of 100kb around sites with significant excessive heterozygosity estimates (F < -0.95 and p value < 10^{-6}) were excluded (Figure S5E). In addition, entire scaffolds where either 20% or more of their total sequence was removed or scaffolds with average F value across all sites F < -0.1 were excluded.

Reference Quality filtering: Summarizing reference quality filters

The union of all masks of the reference genome (scaffold size, mappability, RepeatMasker, aligned depth, autosome identification, excessive heterozygosity), resulted in 1.3 and 0.9 billion accessible bases in the leopard and cat reference genomes, respectively (Table S3), henceforth referred to as *strictref.* The *strictref* filter was applied to all analyses unless otherwise noted.

Sample quality filtering: Error rates estimation

To identify and exclude highly error-prone samples from the dataset, we estimated error rates in ANGSD[21] using the ''perfect-individual'' approach as described in Orlando et al.[72] Sample 3241 was selected as the ''perfect-individual'' sample for the analysis based on the overall mapping statistics. A consensus sequence was generated for this sample with the -doFasta 2 option in ANGSD,[21] taking the most commonly observed base (-doAncError 2) as the consensus. Since the ''perfect-individual'' approach assumes equal genetic distances of the samples to the outgroup, we used the domestic cat reference genome as the ancestral (outgroup) sequence. Error rates were then measured as an excess/deficit of derived alleles from the outgroup compared to the ''perfect-individual'' sample. Both for generating the consensus sequence and for the error estimation, we restricted the data to a base quality of at least 30. Based on the estimated error rates, we removed four samples from the dataset (sample IDs 6350, 6352, 7465, 7466), which differed considerably from the ''perfect-individual'' estimate (Figures S6A and S6B).

Sample quality filtering: BLAST analyses of mtDNA

To ensure that all samples were leopard samples, we investigated the genetic similarity between the mitochondrial genome for each sample and the leopard mitogenome. Mitochondrial genomes for each sample were reconstructed by creating consensus sequences using ANGSD[21] option -doFasta 2 which selected the most common base observed per position on the NC_010641.1 scaffold.[70] We then used BLAST to compare the mitogenomes of our African leopards to the non-redundant BLAST nucleotide sequence database using megaBLAST[64] with a threshold of 90% identity and 75% query cover. All samples were inspected and summary statistics for the samples with extreme results are available in Table S4. Since these overlapped with the samples with high error rates this analysis did not lead to removal of any additional samples.

Sample quality filtering: Identification of related and duplicated samples

Using the methodology described in Waples et al.,[73] we identified and removed closely related and potentially duplicated samples by first inferring two-dimensional site frequency spectrum (2d-SFS) for each pair of samples. Then we calculated three statistics directly from the 2d-SFS for each pair of samples: R0,[73] R1,[73] and KING-robust kinship[74] that can be used to identify close familial relatives[73] without estimates of population allele frequencies.

Using this approach, we identified 12 duplicated samples. All of these had KING-robust kinship values > 0.470, R1 values > 10, and R0 values < 10^{-6}. Within each set of duplicated samples, we removed the sample with the highest error rate, thus removing eight samples for further analyses (sample IDs 6342, 6344, 6348, 6349, 6353, 6556, 6357, 6359). All duplicated samples originated from Zambia (Table S2). We also identified pairs of related samples, down to second-degree relatives (expected kinship = 0.125), based on R0, R1, and KING-robust kinship following the criteria from Manichaikul et al.[74] and Waples et al.[73] In total, we found two pairs of first-degree relatives: one parent-offspring pair and one full-sibling pair, and seven second-degree pairs of relatives (Table S5). For some of the analyses we removed one of the samples from each of the pairs of first-degree relatives (samples 8647 and 7943).

For a summary of the sample sizes used for each analysis, see Table S3.

Genotype likelihoods, SNPs, and genotype calling

We used ANGSD[21] to estimate genotype likelihoods using the genotype likelihood (GL) model from GATK (-gl 2),[65] inferring major and minor from GL data (-doMajorminor 1), estimating allele frequencies from the GL data (-doMaf 1), and applying the *strictref* filter (-sites). We restricted the analysis to bases with base quality of at least 30.

For SNP calling we used a likelihood ratio test (p value = 10^{-6}) as implemented in ANGSD.[21] We only included common alleles (MAF > 0.05).

For the five high-coverage samples, we called genotypes using bcftools (v.1.9).[67] We called the genotypes per sample using the multiallelic caller methodology (-m). We disabled base alignment qualities (BAQ; -B) and restricted the analysis to base qualities of at least 30. In addition to the *strictref* filter (-T), we also excluded sites with a depth of coverage below 10 and heterozygous calls with less than 3 reads support for each allele using plugin setGT to reduce genotype calling errors. These additional genotype calling filters were validated based on heterozygous calls in runs of homozygosity (ROH) in these five samples (Figure S4A).

PCA analyses

We performed PCA (Figure 2A) on 39 individuals based on genotype likelihoods using PCAngsd,[19] which accounts for missingness in low-coverage samples. We excluded the two first-degree relatives (sample IDs 8647 and 7943) from the analysis. We used three eigenvectors to model population structure in the iterative procedure of PCAngsd[19] as anything higher would reflect genetic variation within populations (Figure S2A).

NGSadmix

We estimated admixture proportions for 38 individuals with NGSadmix[20] (Figure 2B) based on genotype likelihoods. Besides removing the two first-degree relatives, we also removed the only individual sampled in Uganda (sample ID 8540), as unbalanced sample sizes prevent from accurately characterizing the genetic composition of its population of origin. We ran NGSadmix from K = 2 to K = 6. For each K we performed several independent optimization runs either until convergence, defined as having the 3 top maximum likelihood results within 2 log-likelihoods unit of each other, or until a limit of 100 runs was reached without convergence. For the values of K where the results converged (Figure S2), we assessed the model fit of the resulting admixture proportion with evalAdmix, which estimates the pairwise correlation of the residual between individuals.[23] Individuals within a population with a bad model fit show a positive correlation of their residuals (Figure 2D).

Runs of homozygosity (ROH) analysis

We estimated ROH with the filtered genotypes of high coverage samples (see STAR Methods). We used the command '–homozyg' in PLINK v1.9[66] with default settings, except using a minimum of 500 kilobases to detect a ROH (–homozyg-kb 500).

D-statistics

We used the *D*-Statistics (ABBA-BABA)[25,26] method to test for population homogeneity and gene flow. We used ANGSD[21] option -doAbbababa 1, which samples a random base at each position to estimate the counts of ABBA and BABA sites between each triplet of H1, H2 and H3 individuals in blocks of a predefined size of 5 MB. We calculated the Z-score based on jackknife procedure[75] with 475 blocks for data mapped to domestic cat reference and 567 blocks for samples mapped to Amur leopard.

First, we explored the homogeneity of the designated groupings (Figure 3). For each sampling location group, we tested all low-depth individuals in a pairwise manner as (H1, H2) against the remaining African groups (H3). We excluded all individuals from Uganda and Tanzania East due to low sample size, and three individuals with distinct ancestry profiles in the NGSadmix analyses with K = 4 compared to the rest of samples in their group (marked with an asterisk in Figure 2). The Amur leopard reference (PanPar1.0) was used as an outgroup (H4). Second, we tested for relative differences in gene flow between the African groups (H1, H2) and the Asian leopard (H3; ID 3042211). As an outgroup, we used the cat reference genome (H4). We excluded one Zambian sample (sample ID 2523) due to low mapping quality, which the *D*-statistics is sensitive to when placed in H1 or H2, and the same three individuals with distinct ancestry profiles in the NGSadmix analyses that were also excluded from the *D*-statistics to test for population homogeneity.

Site frequency spectrum

We estimated the site frequency spectrum (SFS) with the realSFS[28] program (v0.931) within ANGSD,[21,28] using default settings, based on saf files generated with ANGSD (-dosaf 1 -gl 2 -minQ 30). This procedure was used to estimate one-dimensional (1d) SFS per individual and per population and two-dimensional (2d) SFS for all pairs of individuals and populations.

SFS-based analyses: F_{ST} estimation

A genome-wide estimate of F_{ST} (Table 1) for each pair of populations was computed directly from the 2d-SFS, estimated with realSFS,[28] using the Reich estimator.[27] To estimate F_{ST} between a pair of high-coverage individuals, we used as input the 2d-SFS based on called genotypes.

SFS-based analyses: Estimation of heterozygosity

We estimated heterozygosity for each sample in ANGSD[21] as the proportion of heterozygous loci in the obtained per-individual 1d-SFS. For the population and species comparison (Figure 6B), we only used the final dataset of 41 samples, minus two samples with slightly elevated error rates, which show a linear trend between heterozygosity and error rates (Figure S6C).

For genetic diversity comparisons, we obtained whole-genome-based estimates from the literature (see Table S6 for details). To ensure that these were comparable to our estimates we also re-estimated those for the genomes that we re-analyzed for our PSMC analyses.

PSMC

We used PSMC[30] to estimate the population size trajectory back through time for our five high coverage African leopards. We followed the recommended procedure for generating a diploid sequence per individual using bcftools (v 1.9)[67] and -c for calling genotypes. PSMC was applied with default settings. For scaling, we used a mutation rate of 10^{-8} and generation time of 5 and 7.5 years.[5,10,11] Five years is the generation time most commonly used in similar studies, but the IUCN[11] lists the leopard generation time at 7.42 years. Hence, we consider 7.5 more appropriate (Figure 6A), but we also performed the analysis with a generation time of 5 years (Figure S7A) for consistency with previous studies.

For PSMC comparisons, we re-analyzed whole genome data from published studies of other big cats, including Amur leopards,[5] tiger,[31] lion,[31] jaguar,[10] and snow leopard.[31] Briefly, we mapped the sequencing data to the cat reference genome using the same pipeline as for the African leopard sequencing data. Unless otherwise noted, we also applied the *strictref* filter to all analyses.

Identity-by-state (IBS)

We estimated the pairwise genetic distance matrix (IBS) between all individuals by randomly sampling a single read from each position from each sample applying the *strictref* filter, using -doIBS in ANGSD. This determines the allelic distance (0 or 1) in sites with information from both individuals and calculates the average identity-by-state (IBS) across all sites.

IBS-based analyses: Neighbor-joining (NJ) tree

We used the IBS matrix to construct a NJ tree (Figure 2C) using the R package ape[22] (https://cran.r-project.org/package=ape). The NJ tree included samples with distinct ancestry profiles as inferred from NGSadmix, denoted with an asterisk.

IBS-based analyses: Estimation of Effective Migration Surfaces (EEMS)

We used the pairwise IBS matrix as input to EEMS[29] to estimate the relative migration rates (Figure 5, Figure S3). For each sampling locality we visually inferred the centroid of its geographical coordinates and used this as the coordinate input to EEMS. EEMS was run using the runeems_snps program and default settings for 5 million steps and a burn-in of 1 million steps, with 400 demes and the number of sites (n = 1,374,592,127) based on the IBS matrix. For comparison, we ran EEMS with the same settings for other African mammals with available SNP datasets - waterbuck,[19] African buffalo (Nursyifa et al., in prep.), and common warthog (Jørgensen et al., in prep.). Furthermore, we included the savanna African elephant[76] and African lion,[77] for which microsatellite datasets with continent-wide sampling were available. For the microsatellite data, the runeems-sats was run using the genotypes and individual sample coordinates as input and the same settings as above. The elephant data by Wasser et al.[76] has been previously used for the estimation of relative migration rates in the original EEMS publication.[29] Unlike Petkova et al. we used the full savannah elephant dataset (n = 1001), including samples assumed to be hybrids; however, the results are very similar.

For all species, the same habitat outline was used and was selected to encompass the sampling localities of all included species. The outline was drawn manually using the tool available here: http://www.birdtheme.org/useful/v3tool.html. The results were visualized in RStudio v1.2.5033 using the make_eems_plots script (available here: https://github.com/dipetkov/reemsplots2) as the log10-transformed migration rates with posterior probabilities > 0.90. An outline of the equatorial rainforest was added based on the spatial distribution of African rainforests in Mayaux et al.[18]