Selection against admixture and gene regulatory divergence in a long-term primate field study

Tauras P. Vilgalys*, Arielle S. Fogel*, Jordan A. Anderson, Raphael S. Mututua, J. Kinyua Warutere, I. Long'ida Siodi, Sang Yoon Kim, Tawni N. Voyles, Jacqueline A. Robinson, Jeffrey D. Wall, Elizabeth A. Archie, Susan C. Alberts, Jenny Tung

corresponding author: Jenny Tung, jt5@duke.edu
* These authors contributed equally to this work

This PDF file includes:

> Materials and Methods
> Figs. S1 to S8

Other Supplementary Materials for this manuscript include the following:

> Tables S1 to S9 (Excel)
> MDAR Reproducibility Checklist (PDF)

Materials and Methods

1. Baboons in this study

1.1. Baboon taxonomy, evolutionary history, and divergence

Members of the genus *Papio* are large-bodied, primarily terrestrial cercopithecine monkeys with origins in southern Africa. They subsequently expanded across the continent, forming two distinct lineages—the northern and southern baboon clades—approximately 1.4 million years ago, albeit with multiple lines of evidence for gene flow between lineages after this date (*14-16, 48*). The current scientific consensus recognizes six extant species: the chacma baboon (*Papio ursinus*), the Kinda baboon (*P. kindae*), and the yellow baboon (*P. cynocephalus*) in the southern clade, and the anubis (or olive) baboon (*P. anubis*), the hamadryas baboon (*P. hamadryas*), and the Guinea baboon (*P. papio*) in the northern clade (*14, 48*). Thus, the two taxa that are the focus of this study, the anubis baboon and the yellow baboon, belong to the northern and southern lineages, respectively, and are therefore the mostly distantly related that extant species in *Papio* can be.

We note, however, that baboon taxonomy and phylogeny have long been debated, with the genus split into anywhere from one to sixteen species (*14*). The current six-species classification is based on stereotyped phenotypic differences between baboon taxa, their largely non-overlapping geographic ranges, and recent phylogenetic reconstructions based on whole-genome sequencing data (*14*). Regardless, like many other primates that hybridize in nature (reviewed in (*4, 5, 49, 50*)), baboons are not good "biological species": they produce viable and fertile hybrids where their geographic ranges meet ((*8, 51, 52*); reviewed in (*12*)), and their evolutionary history includes multiple bouts of documented or inferred cases of divergence followed by reconvergence (possibly, in one case of localized northern-southern clade reconvergence, giving rise to the modern Kinda baboon, and in some cases involving ghost lineages that can no longer be sampled) (*14, 16*). While this history complicates taxonomy, it also makes baboons a useful model for the complex evolutionary history that similarly characterizes the human lineage. To minimize confusion, and in alignment with other recent literature on baboon genetic and phenotypic variation (*12, 14*), we therefore follow the current convention for baboon taxonomy, which confers species status to six species, including the taxa of interest in this analysis.

1.2. The Amboseli baboon population

This study primarily focuses on a wild population of baboons inhabiting the Amboseli ecosystem of southern Kenya. Individuals from this population have been subject to near daily monitoring by the Amboseli Baboon Research Project (ABRP) since 1971 (*17*). When long-term observation began, the population was thought to include only yellow baboons (*33*), but immigrant anubis baboons and yellow-anubis hybrids have been documented since 1982 (*18, 34, 51*).

Within the present sample of Amboseli individuals (n=442), 42.5% are known descendants of these anubis/anubis-like immigrants based on pedigree evidence (hereafter, referred to as "recently admixed" individuals, n=188). As suggested previously and confirmed here, however, even those individuals with no recent anubis ancestry appear to be admixed from events that predate long-term monitoring (hereafter,

referred to as "historically admixed" individuals; n=214) (*15*). Recently admixed individuals therefore contain anubis ancestry from both recent and historical admixture events. For the 40 remaining Amboseli animals, classification into these two categories was ambiguous because animals had no recent anubis ancestor but elevated levels of anubis ancestry consistent with an undetected, recent anubis ancestor (mean ancestry ≥ 30%, n=26) or because we were not completely confident about the individual's identity (n=14, see Section 5).

In cases in which we analyzed the genomes of recently admixed and historically admixed individuals separately, we focused on chromosomes with ≥ 30% estimated anubis ancestry for the recently admixed group. We did so because anubis ancestry is distributed unevenly across the genomes for recently admixed individuals, who are products of historical admixture as well. Consequently, some chromosomes in recently admixed individuals primarily reflect historical, not recent, admixture. Despite this step, we note that the signatures we observe in recently admixed individuals are likely driven in part by historical ancestry that co-exists on chromosomes with recent anubis ancestry.

Permission for research on baboons in Amboseli was granted, at various points in the long-term study's history, by the Kenya Wildlife Service, the CITES-granting authority in Kenya; the National Environmental Management Authority, which issues permits for biological sample collection; and the National Commission for Science, Technology and Innovation, which issues permits for conducting research. We currently obtain or renew multiple permits from each of these bodies on an annual basis, but note that this procedure has differed over time because the responsible bodies in Kenya and the permitting requirements have also shifted (e.g., research authority over scientists working in the Kenyan national parks shifted to a new body, the Wildlife Research & Training Institute, in 2021). Pursuant to Kenya's signing of the Nagoya Protocol (*53*), the ABRP has also entered into a Memorandum of Agreement (MoA) that includes the US-based institutions that host our research, the Kenya Wildlife Service, the University of Nairobi, the Nairobi-based Institute of Primate Research, and the National Museums of Kenya. This MoA details the ABRP research mission and commitment to benefits sharing under the Nagoya Protocol, which include student training at all educational levels; employment of Kenyan nationals; collaborations with Kenyan researchers; and sharing of biological samples and data.

1.3. Other anubis and yellow baboon populations
To characterize the genetic structure of the yellow-anubis hybrid zone, we also analyzed individuals from two putatively unadmixed yellow baboon populations and four putatively unadmixed anubis baboon populations. Yellow baboons were sampled from (i) Mikumi National Park in central Tanzania (n=15); and (ii) the wild-caught founders of the Southwest National Primate Research Center (SNPRC, now part of the Texas Biomedical Research Institute), originally captured in Kenya in 1972 (n=7). Anubis baboons were sampled from (i) the Aberdares range in central Kenya (n=2); (ii) the Maasai Mara National Reserve in Kenya (n=7); (iii) the wild-caught founders of the SNPRC colony (n=24); (iv) two descendants of those founders, also at the SNPRC; and (v) the Washington National Primate Research Center (WNPRC; n=6). Because the trapping locations for the SNPRC founders were close to the known hybrid zone (*15*) (Fig. 1A), it is possible that the colony founders may have been admixed themselves.

This possibility was supported in our analyses for the yellow baboon founders, leading us to mask potentially introgressed regions when calculating reference population allele frequencies (Fig. 1; Section 6; Section 9). All other populations are far from the hybrid zone and unlikely to be admixed based on genetic data, phenotypic data, and/or field reports (*14, 15, 22, 54, 55*).

2. Sample collection

In this study, we used a combination of previously published and newly generated genomic data. Previously published whole-genome resequencing data were available for baboons from Amboseli (n=22), the Aberdares (n=2), Maasai Mara (n=7), Mikumi (n=5), the SNPRC (n=33), and the WNPRC (n=6) (*14, 15, 23, 24*). Previously published RNA-seq data were available for 145 unique individuals (157 samples) from Amboseli (*40-42*).

We also generated new whole-genome resequencing from individuals in Amboseli (n=420) using blood or other tissue samples collected by the Amboseli Baboon Research Project, for a total of 442 sequenced Amboseli baboons. The majority of samples were blood-derived, collected from animals that were anesthetized with a Telazol-loaded dart using a handheld blowgun. This procedure follows a highly conservative protocol focused on maintaining habituation and minimizing the possibility of injury (*56, 57*). Sample storage for DNA extraction varied over time because blood samples were collected from 1989-2018. However, in all cases, blood or purified white blood cell samples were placed in a storage medium (either a Tris-based storage buffer or RNALater) and archived at -80° C upon arrival in the United States. Samples for RNA extraction were collected in either PaxGene RNA tubes (Qiagen; for samples collected from 2007-2010 (*40*)) or TruCulture "null" (media-only) tubes (Myriad RBM; for samples collected from 2013-2018 (*41, 42*)). Following sample collection, animals were placed in a covered holding cage to recover from the effects of the anesthesia and then released to their social group under observer monitoring. For an additional 32 samples used to generate resequencing data, we opportunistically collected non-blood tissue samples (e.g., from recovered corpses), stored in either 95% ethanol or RNALater until DNA extraction. An additional 5 samples were obtained from previously extracted DNA where the original sample was unavailable, but was almost certain to be blood collected during darting.

3. Data generation

We sequenced DNA from 430 baboon samples. For 423 individuals (n=413 Amboseli animals, 10 Mikumi animals), we prepared DNA sequencing libraries using the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs). We input 5 ng of DNA per library, and used half-reactions for fragmentation/end repair, adapter ligation, and PCR enrichment. Libraries were dual-indexed for sequencing using NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Sets 1 and 2) and amplified for 8 PCR cycles. We assessed library size and quality on a 2100 Bioanalyzer System (Agilent). For seven additional Amboseli individuals, DNA sequencing libraries were prepared by the Garvan Institute of Medical Research in Australia.

Libraries were sequenced on the Illumina HiSeq X Ten at MedGenome, Inc. or at the Garvan Institute. We generated moderate to high coverage resequencing data for seven Amboseli individuals (mean coverage from non-read duplicate reads mapped at

MAPQ ≥ 10 = 35.56 ± 3.81 s.d.) and 10 yellow baboons from Mikumi (mean coverage = 13.38 ± 1.50 s.d.), and low coverage resequencing data for the remaining 413 individuals from Amboseli (mean coverage = 1.04 ± 0.19 s.d.) (Table S1). We combined data from Amboseli with previously generated resequencing data from an additional 22 animals (mean coverage = 3.60 ± 6.58 s.d.) (*15, 23*) to generate a data set of 442 individuals from the Amboseli population (Table S1). We also combined the Mikumi data with previously generated resequencing data from 51 yellow or anubis baboons (mean coverage = 24.83 ± 18.42 s.d.) (*14, 15, 24*) to generate a data set of 61 baboons from other yellow and anubis baboon populations (Section 1.3; Table S1).

4. Mapping and variant calling

We generated joint variant call format (VCF) files using a pipeline adapted from the Genome Analysis Toolkit (GATK (*58-61*)) Best Practices (*62*). First, we used *TrimGalore* (*63*) to remove adapters and low-quality ends of reads (minimum Phred score threshold of 20) and imposed a minimum read length threshold of 50. Each sample was then processed individually to generate gVCF files as follows. We used *Picard*'s (*64*) FastqToSam to convert fastq files to BAM format and add unique read group names. We aligned reads from all samples to the anubis baboon reference genome, *Panubis1.0* (*65*), using *bwa-mem* (*66*). Specifically, we used piped commands for this step: Picard SamToFastq | bwa mem | Picard MergeBamAlignment. While mapping to a heterospecific reference genome can cause mapping biases, we found no significant differences in the percentage of reads mapping to the genome by ancestry (SNPRC anubis baboon founders vs. SNPRC yellow baboon founders: Mann-Whitney test p = 0.44; SNPRC anubis baboon founders vs. Mikumi yellow baboons: p = 0.97; mapping results for all samples are provided in Table S1).

Next, we tagged PCR duplicates with *Picard*'s MarkDuplicates function. We then generated per-sample gVCF files with GATK *HaplotypeCaller*, using bases with minimum base quality scores of 10 and reads with minimum mapping quality scores of 20. Next, we performed joint genotyping using batches of gVCF files produced by GATK *CombineGVCFs*. Joint genotyping was performed with GATK *GenotypeGVCFs*. The resulting VCF files were then left-aligned using GATK *LeftAlignAndTrimVariants*, and alternate alleles that did not appear in any samples were removed with the *trimAlternates* option of the GATK *SelectVariants* tool.

We then filtered for high quality variants following GATK's recommended criteria for hard filtering for germline variants. Specifically, we retained biallelic variants with QualityByDepth (QD) > 2, Fisher strand bias (FS) < 60, root mean square mapping quality (MQ) > 40, mapping quality rank-sum test (MQRankSum) > -12.5, and rank sum test (ReadPosRankSum) > -8. To call a variant within the Amboseli baboon population, we also required that the position be genotyped in at least 20% of individuals and be genotyped in at least half of the anubis and yellow baboons used in our reference panels. In the final, post-filtering set of variants, the minimum, maximum, and mean across Amboseli samples for mean allele depth (DepthPerAlleleBySample i.e., the AD field) were 0.399 (min), 0.846 (mean), and 2.507 (max) for reference alleles in low coverage samples and 0.056 (min), 0.217 (mean), and 0.836 (max) for alternate alleles in low coverage samples (<4x; n=433). For high coverage samples (>10x; n=9), the

corresponding values were 16.29 (min), 24.11 (mean), 29.56 (max) for reference alleles, and 5.92 (min), 8.89 (mean), and 10.96 (max) for alternate alleles.

5. Resequencing data set: quality control for individual identity

For the Amboseli data set, we confirmed that our genetic data matched their donor labels based on correct assignment of sex and reconstruction of kin relationships consistent with the independently generated, multigenerational Amboseli pedigree. For cases in which RNA-seq data were also available, we also confirmed that genotypes from the resequencing data and the RNA-seq data matched within individuals.

For sex, we checked whether resequencing data from females had approximately 2x the proportion of mapped reads assigned to the X chromosome relative to resequencing data from males (Fig. S1A). For pedigree relatedness, we compared pedigree-based relatedness calculated using the R package *pedantics* (*67*) (version 1.7) to genetic estimates of relatedness based on the resequencing data, calculated using *lcMLkin* (*68*) (designed for low-coverage data). Pedigree maternal links are based on field observations of pregnancy and birth. Pedigree paternal links are based off demographic data (e.g., the presence of reproductively mature males in a female's social group at the time of conception) and microsatellite genotyping followed by paternity assignment by exclusion (*69, 70*) (extragroup paternity has never been observed in this population).

For each individual, we plotted their pedigree versus resequencing-based estimates of relatedness with all other individuals in the data set, paying particular attention to concordant estimates for known close kin dyads (e.g., parent-offspring, full sibling, half-sibling, grandparent-grandoffspring) (Fig. S1B). Out of 442 individuals, we identified six cases where the sample did not match the original labeled individual and a further six individuals whose identity could not be unambiguously confirmed by the low-coverage data. In addition, two individuals were sampled from non-study groups (i.e., baboon groups that also live in the Amboseli ecosystem but are not regularly monitored to collect behavioral and demographic data). We removed these 14 individuals from analyses that depended on assignment of individual identity (i.e., because they required phenotypic or demographic data), but retained them for analyses that did not require individual-specific information, as we could confirm they were distinct from the other individuals in our sample.

For individuals with paired genome resequencing and RNA-seq data, we calculated the correlation between genotypes called from the resequencing data against genotypes called from all RNA-seq samples. In all cases, the highest resequencing-RNA-seq correlation was observed between samples from the same individual.

6. Local ancestry calling

6.1. Statistical method

We estimated local ancestry for each individual in the resequencing data set using LCLAE (**L**ow **C**overage **L**ocal **A**ncestry **E**stimation) (*15*). This method estimates the number of introgressed alleles at each SNP based on the genotype likelihoods for nearby sites (here, 0 = homozygous yellow; 1 = heterozygous; 2 = homozygous anubis). LCLAE thus draws on allele frequency differences and uncertain genotype calls integrated across variants to reliably assign ancestry from low coverage data. Specifically, LCLAE

calculates the likelihood of each possible ancestry state Y_i, where Y_i is 0, 1, or 2 and corresponds to the number of alleles of introgressed ancestry (here, anubis). This likelihood is given by:

$$lik(Y_i|G) = \sum_{j=0}^{2} \Pr\left(G_j\right) lik\left(Y_i|G_j\right)$$

where G is the genotype (a set of three possible values) and j refers to the number of alternate alleles. To combine information across SNPs, a composite likelihood is calculated by multiplying the probabilities of observed genotypes across sites in a predefined window size. The ancestry state with the greatest composite likelihood score is the one assigned for that SNP; in the case of ties, no ancestry is assigned. Then, for each SNP, a final ancestry call state is made using majority rule on all windows overlapping that SNP.

6.2. Reference panels

LCLAE requires allele frequency information for the parent taxa for inference. To generate this information, we used allele frequencies from the 7 SNPRC yellow baboon founders and the 24 SNPRC anubis baboon founders, all of whom were sequenced to high coverage (*24*). We masked any regions of the genome that might be affected by introgression, as our analyses suggest that the SNPRC yellow baboon founders likely harbor low levels of anubis baboon ancestry from past admixture events (Fig. 1; Section 9; note that we masked only those regions in the specific individuals for whom anubis ancestry was detected, not in our analysis as a whole).

To identify putative introgressed regions to mask for each reference panel individual, we used a conservative intersection set of loci identified as (i) introgressed based on local ancestry calls with LCLAE, using reference panel allele frequencies calculated without that individual included; and (ii) identity-by-descent (IBD) with members of the reference panel for the other species using IBDmix (*26*). For IBDmix, which performs pairwise analyses, we required consensus IBD calls from at least 50% of reference animals from the other species. Nearly all potentially introgressed regions identified by LCLAE are included in the set of IBD calls identified by IBDmix (Section 9.2). Our approach masks 14.9-34.2% of SNPRC yellow founder genomes and 2.0-6.6% of SNPRC anubis founder genomes. However, because tracts were masked on a per individual basis rather than across the entire reference population, and introgressed segments differ across individuals, we were able to obtain ancestry calls for 99.9% of the genome in both reference populations. In addition, we found that, while masking improves ancestry calling consistency within Amboseli trios (see Section 7.3), it has only a minor impact on ancestry calls: $92.6 \pm 0.98\%$ s.d. of local ancestry calls for Amboseli samples are identical when using a masked vs. unmasked SNPRC panel. However, masking increases the density of ancestry informative loci (i.e., SNPs with large allele frequency differences between yellow and anubis baboons) across the genome relative to unmasked genomes.

6.3. Local ancestry tracts

We performed local ancestry analysis for all study subjects in all populations, including reference population animals after removing the focal animal from the

reference panel before inference. To do so, we focused on ancestry informative markers that exhibited at least a 20% difference in allele frequency between the yellow and anubis reference populations and that were also genotyped in Amboseli. We calculated composite likelihood values for 35 kb windows centered on each ancestry informative marker and then estimated local ancestry at that marker using majority rule for sites within 35 kb upstream or downstream of the focal marker (i.e., ancestry was assigned based on the majority ancestry state when ≥50% of sites within the 70 kb centered on each site were assigned the same state). We removed ancestry calls from windows with fewer than 20 nearby ancestry informative markers, as low marker density contributes to errors in ancestry calling. These parameters and window sizes were selected based on simulations to maximize accurate ancestry assignment (Section 7.1). Finally, we collapsed sequential markers that were assigned the same ancestry state into contiguous tracts. We placed tract break points exactly halfway between sequential SNPs in which ancestry state assignments switched. We also removed ancestry tracts shorter than 1 kb, as simulations indicate that tracts of this size are unreliable (Section 7.1).

7. <u>Accuracy</u> of <u>local</u> <u>ancestry</u> <u>assignment</u> and <u>comparisons</u> to <u>alternative</u> <u>methods</u>

7.1. Simulated data and LCLAE optimization

To evaluate the accuracy of LCLAE local ancestry calls, we simulated 25 generations of admixture for a population of 1,000 baboons with 30% anubis ancestry (approximately the mean ancestry in the current Amboseli population), discrete generations, limited immigration (10 individuals per generation), and a uniform recombination rate using SELAM (*71*). Simulated individuals mated randomly and no variants were under selection. This simulation produced "true" local ancestry tracts for 25 individuals with a length distribution similar to those inferred for real data from 9 Amboseli baboons sequenced at high coverage. Because simulated ancestry tracts are known (i.e., they are "ground truth"), we can assess the performance of local ancestry calling algorithms by comparing the percentages of simulated genomes that are assigned the "true" ancestry state.

Variant-level genotypes for each simulated individual were drawn based on the simulated individual's known ancestry at each ancestry informative marker and reference panel allele frequencies for yellow and anubis baboons. For these simulations, we used the same reference panels that we applied in real data analysis (7 yellow baboon and 24 anubis baboons from the SNPRC founders). We then simulated sequencing reads (10x coverage) for each individual using the program NEAT-genReads (*72*). After mapping simulated reads using *bowtie2* (*73*) and calling sample genotypes as above, we called local ancestry for each simulated individual using LCLAE and the same SNPRC founder reference panels. To assess the effects of sequencing depth, we also down-sampled the mapped bam files to 1x coverage (similar to most of our resequencing data for Amboseli), re-called genotypes, and re-estimated local ancestry. To optimize the set of ancestry informative sites and window sizes used to calculate composite likelihoods, we also explored the relative accuracy of local ancestry calls when setting the minimum yellow-anubis allele frequency difference for ancestry informative sites to values between 5 and 50%, and using ancestry informative sites in window sizes of 10 to 50 kb (with windows twice these sizes used to smooth ancestry calls using majority rule).

To assess local ancestry calling accuracy, we calculated the percentage of the entire genome where the estimated ancestry matched the simulated ("true") ancestry. LCLAE performs very well for both high and low coverage data using 35 kb windows, correctly assigning between 95 and 99% of the genome to the appropriate ancestry state (Fig. S2A; mean accuracy = $97.9 \pm 0.4\%$ s.d. for high coverage sequencing and $96.9 \pm 0.6\%$ for 1x coverage sequencing). About half of the errors in calling local ancestry in the low coverage data were also errors in the high coverage data, suggesting that errors are often due to LCLAE's assumptions themselves, not to low coverage.

Sequencing coverage and window size interact to influence the accuracy of ancestry calls. There is a trade-off between the number of ancestry informative sites used to make ancestry calls—there are more sites in larger windows and in high coverage data—and the possibility of missing small ancestry tracts because larger windows can span multiple smaller tracts. For low (1x) coverage data, a 35 kb window size maximized the proportion of the genome that was assigned to the "true" ancestry state in our simulations (96.9%; Fig. S2B). High coverage data followed a similar pattern, but tended to produce reasonably accurate estimates even with smaller window sizes, presumably due to a greater density of typed ancestry informative sites per sample. Because most of our sequencing data for Amboseli is low coverage, we used 35 kb windows in our subsequent analyses. We note that local ancestry calling is highly accurate for longer ancestry tracts (reaching 95% accuracy for tracts greater than 50 kb, even in low coverage data) but highly error-prone for tracts shorter than 1 kb (only 40% accuracy for high coverage data) (Fig. S2C-D). We therefore eliminated any tract calls less than 1 kb. We found no appreciable difference in the accuracy of ancestry assignments with minimum allele frequency differences between 5 and 30%. Accuracy decreases when using a higher threshold for yellow-anubis allele frequency differences (e.g., 50%) because of the sparsity of highly divergent sites with genotype calls, particularly in low coverage data. However, small differences in allele frequency are likely to be uninformative regarding ancestry in real data because estimated allele frequencies are drawn with sampling error and may not accurately reflect parental populations. We therefore chose a threshold of 20% minimum allele frequency difference between yellow and anubis baboons, which produces ancestry calls that are robust to sampling error in the reference populations (see below) and likely to represent true differences between species.

We also assessed the robustness of our approach to errors in estimating reference panel allele frequencies (a concern because our panels were relatively small). To do so, we treated the estimated anubis and yellow baboon allele frequencies from the reference panel as "true" values, sampled genotypes based on these frequencies, and simulated reads for 10 yellow baboon and 10 anubis baboon genomes containing the sampled genotypes. Note that because we sampled genotypes from all common variants, the set of highly differentiated variants used for ancestry calling is different in the simulated data relative to the observed data. 78% of the ancestry informative sites identified in the original, observed data are also selected as ancestry informative sites in the simulated data (i.e., exhibit at least a 20% allele frequency difference between resampled, simulated anubis and yellow individuals). Of ancestry informative sites selected in the simulated data, 86.6% were also differentiated by at least a 20% yellow-anubis allele frequency difference in the original, observed data. Most importantly, substituting the allele frequencies estimated from these simulated reference panels for the original, observed

allele frequencies had little effect on the accuracy of local ancestry calling. Ancestry calls using LCLAE for low coverage data still called the true, simulated ancestry state for 94.5% of the genome (compared to 96.9% when using the original, observed allele frequency estimates).

Finally, although our simulations support high accuracy of ancestry assignments in low coverage data, we noticed that in our Mikumi sample, high coverage samples were assigned lower levels of anubis ancestry overall than low coverage samples from the same population (Section 9.2). We therefore directly assessed the impact of coverage on ancestry assignment by downsampling high coverage Mikumi samples to ~1x coverage. Consistent with our simulations, agreement between high and low coverage data from the same individuals was high (~94%). However, we detected on average ~6.5% ($\pm$2.1% s.d; range: 3.9-11.2%) more anubis ancestry at 1x coverage than we did in the original, high coverage data for the same individuals. Because true anubis ancestry in the Mikumi yellow population is near zero, we note that errors due to sequencing coverage will be biased towards increasing, rather than decreasing, estimated anubis ancestry estimates (in contrast, in highly admixed populations like Amboseli, our simulations suggest that errors likely occur in both directions). Thus, although we provide global ancestry ranges in the main text based on this 6.5% estimate, it likely represents an upper bound for biased anubis ancestry inferred due to sample coverage.

7.2. *Comparison* *to* *other* *methods*

In addition to estimating local ancestry using LCLAE, we tested two alternative approaches to calling local ancestry with low coverage data: Ancestry HMM (*74*) and AD-LIBS (*75*). Whereas LCLAE uses a composite-likelihood method to call local ancestry, the other two approaches use Hidden Markov Models to infer the location of ancestry tract breaks. Further, while LCLAE works with genotype likelihoods, Ancestry HMM uses read counts from genotype calls and AD-LIBS uses mapped reads to create a pseudo-haploid genome per individual (a genome where each position is drawn at random from among mapped reads, if any are available). As for LCLAE, we focused on ancestry informative sites with a minimum allele frequency difference >20% between yellow and anubis baboons and placed tract breaks exactly in the middle between variants that were assigned different ancestry states. In these approaches, there is no analog to the window size parameter for LCLAE. For Ancestry HMM, we assumed 30% anubis ancestry and two pulses of anubis introgression, which were allowed to vary in time and degree of gene flow between species. For AD-LIBS, we also assumed 30% anubis ancestry.

For 10x coverage data, the Ancestry HMM approach performed best, matching the simulated ancestry for more than 99% of the genome (mean accuracy = 99.7 $\pm$ 0.1% s.d.; Fig. S2A). LCLAE performed slightly worse on 10x coverage (97.9 $\pm$ 1.0%). However, at low-coverage (1x), LCLAE retained high accuracy (96.9 $\pm$ 1.1%), whereas Ancestry HMM's accuracy dropped to 88.4 $\pm$ 3.8%. AD-LIBS performed surprisingly poorly at both 1x and 10x coverage, only assigning the correct ancestry 40 $\pm$ 5.6% of the time (Fig. S2A). This is likely because AD-LIBS uses only one mapped read per position, making it difficult to accurately assign ancestry between taxa that share substantial amounts of genetic variation (such as yellow and anubis baboons), as opposed to taxa with many more fixed differences (e.g., polar and brown bears, the admixture case that motivated the

original development of AD-LIBS (*75*)). Because most of our data set for this study is low-coverage data, we therefore chose to use LCLAE to assign local ancestry.

7.3. Consistency of ancestry calls within pedigree trios
We also evaluated the quality of our local ancestry calls using known parent-offspring trios from the Amboseli population. As a result of long-term monitoring, maternal-offspring links are known based on near-daily observations of pregnancies and the appearance of neonates; paternities have been assigned for the past several decades based on multilocus microsatellite genotyping (*69, 70*). Our resequencing data set contained 92 unique mother-father-offspring trios (n=181 unique low coverage individuals), in which, at any given locus, only a subset of possible local ancestry state calls would be consistent with the trio structure (Fig. S3A-C). We assessed consistency of local ancestry calls based on 73,975 loci across the genome that were spaced at least 35 kb apart and at least 50 kb from the ends of chromosomes. For each trio, we also removed loci that fell in tracts <1 kb length or that fell on the exact base pair that transitioned between ancestry states. For each trio at each remaining site, we evaluated whether the inferred ancestry of the offspring was consistent with the inferred ancestries of the parents (hereafter, "pedigree consistencies"). We then calculated the per-site proportion of pedigree consistencies across all trios with data available for all three individuals.

Compared to the low-coverage reference panel used by Wall et al. (*15*), supplemented with 11 higher coverage Mikumi samples (5 of which were higher coverage data from individuals already sequenced by Wall et al. (*15*)), pedigree consistency was higher when using high-coverage, masked reference panels from the SNPRC anubis and yellow baboon founders (Fig. S3D; median proportion of pedigree consistencies using the SNPRC panel = 0.924 ± 0.062 s.d. vs. 0.717 ± 0.099 s.d. using the supplemented Wall et al. panel). Across loci, we achieved greater consistency when the 35 kb windows centered on each site (i) contained more ancestry informative markers ($p<10^{-300}$); (ii) were more diverged between yellow and anubis baboons ($p<10^{-300}$; based on weighted F_{ST} calculated using *vcftools* (*76*) [version 0.1.15]); and (iii) were in regions of elevated recombination ($p<10^{-32}$; mean recombination rates were $\log_{10}$ transformed and calculated for 250 kb windows centered on each site instead of 35 kb windows because recombination rates from 35 kb windows are likely to be estimated with more noise (see Section 12.3 for details on recombination rate estimates; Table S7). Across trios, consistency was lower when one member of the trio was sequenced at particularly low coverage (Pearson's r between the minimum genome-wide coverage among members of the trio and the proportion of sites with consistent calls within trios = 0.553, p = 1.07×10^{-8}).

7.4. Ancestry calls are robust to choice of reference population
The ancestral populations that best represent the yellow and anubis fractions of the genome in the Amboseli baboons are not known. Further, we chose to use wild-caught founder individuals for the SNPRC population as our reference populations, meaning that the allele frequencies we used to represent yellow and anubis baboons may reflect biases introduced by capturing/sampling those founder individuals, rather than the "true" ancestral populations. This potential problem is more acute for the yellow baboon

reference population than the anubis baboon reference population, given the smaller sample size available for yellow baboon reference individuals (n=7 versus n=24).

We therefore asked whether our local ancestry calls remained stable if we substituted data from yellow baboons sampled in Mikumi National Park, Tanzania (n=11 genomes sequenced to moderate to high coverage) for the SNPRC yellow founders. If the yellow baboon allele frequencies estimated for SNPRC are not representative, this switch should alter our ancestry calls in Amboseli. We found that ancestry calls are not greatly affected by the yellow reference population we used. Global ancestry for the 442 Amboseli animals is strongly correlated (r = 0.979) between the two analyses and a mean of 82% (± 2.5% s.d.) of the genome is assigned the same ancestry state. Further, the evidence for selection against introgression remains robust using this alternative yellow baboon reference population (Table S3, rows labeled "LCLAE: SNPRC anubis, Mikumi yellow; masked").

For anubis baboons, we did not have another set of high coverage anubis samples that we could substitute for the SNPRC anubis founders as the anubis reference population. However, allele frequencies were similar between the SNPRC anubis founders and the low coverage Maasai Mara anubis population (r = 0.761 and r = 0.822 for ancestry informative SNPs), suggesting that ancestry calls would also be consistent using an alternative reference population.

8. F_4 statistics

In addition to estimating admixture using a local ancestry approach, we used F_4-ratio estimation (25, 77-79) as an orthogonal method to estimate the anubis ancestry proportions of Amboseli individuals. Following the nomenclature of Patterson et al. (25), F_4-ratios estimate the ancestry proportion of members of population X as a combination of two sources, populations B' and C'. Because B' and C' are unknown ancestral populations, they are represented by modern populations B and C. This method also requires two additional populations: population A, which has not contributed to population X but is a sister group to population B, and an outgroup, O, to populations A, B, and C. The contribution of population B' to population X is α (where 1-α therefore corresponds to the contribution of population C' to population X) and $\hat{\alpha} = \frac{f_4(A,O;X,C)}{f_4(A,O;B,C)}$ (see Patterson et al. (25) for details).

To estimate F_4-ratios, we focused on Amboseli animals for which we had high coverage data (n=9), which include two recently admixed animals and seven historically admixed animals. Amboseli is therefore population X. We also used data from the SNPRC anubis founders (population B; n=24) (24), Mikumi National Park in Tanzania (population C; n=10), hamadryas baboons and Guinea baboons resequenced by the Baboon Genome Sequencing Consortium (14) (as two alternatives for population A; n=2 in each case), and either the gelada monkey (also from the Baboon Genome Sequencing Consortium) or rhesus macaque as the outgroup O (n=1 in each case). We mapped these data to the rhesus macaque genome (*MacaM*) (80) using *bowtie2* (73). Duplicate reads were removed using the *MarkDuplicates* function from *Picard* (64) and reads with MAPQ less than 10 were removed using *samtools view* (81). gVCF files were generated using GATK *HaplotypeCaller* (61) requiring a minimum base quality ≥ 20 and then merged across individuals using *CombineGVCFs*. Genotypes were then called using *GenotypeGVCFs*. We filtered for high quality variants following GATK's recommended

criteria for hard filtering for germline variants (see Section 4) and retained biallelic SNPs that were typed within all individuals in the sample. We removed indels, singleton and doubleton variants, and removed clusters of 3 or more variants that fell within a 10 bp window, as dense clusters of SNPs are enriched for false calls (*82*). Since we used the macaque as one possible outgroup, we added the macaque genotype (homozygous reference) as an additional sample at all sites.

 We next converted our calls, in vcf format, to PLINK format (using *vcftools* (*76*), *bcftools* (*83*), and *PLINK* (*84-86*) [version v1.90b3.36]) and then to EIGENSTRAT format (using *convertf* (*25*) in the *EIGENSOFT* package [version 6.1.4]). To calculate the F_4-ratio, we input the EIGENSTRAT formatted data into *admixr* (*87*) (version 0.9.1), an R package that builds on the *ADMIXTOOLS* software suite (*25*). Using the *f4ratio* function, we generated estimates of α per chromosome separately for recently and historically admixed Amboseli animals (i.e., we treated them as two separate *X* populations) for the following phylogenetic configurations (*A*, *B*, *C*, *O*): (i) hamadryas, SNPRC anubis, Mikumi yellow, macaque; (ii) Guinea, SNPRC anubis, Mikumi yellow, macaque; (iii) hamadryas, SNPRC anubis, Mikumi yellow, gelada; and (iv) Guinea, SNPRC anubis, Mikumi yellow, gelada. To obtain a genome-wide estimate of α, we first averaged α estimates for the different phylogenetic configurations per chromosome. Estimates of α were highly stable across phylogenetic configurations (range of s.d. across phylogenetic configurations per chromosome = 0.004-0.037). We then averaged values of α across chromosomes, weighted by chromosome length. Using this method, the mean genome-wide proportion of recently admixed Amboseli genomes derived from anubis gene flow is 0.508 (versus 0.511 from LCLAE) while for historically admixed animals it is 0.232 (versus 0.249 from LCLAE).

9. Assessment of admixture outside of Amboseli

9.1. Principal components analysis and IBDmix

 Yellow-anubis admixture is well-documented within the Amboseli population, but may also occur outside of the known hybrid zone (as suggested by identification of an admixed animal from the Aberdares region of Kenya (*14*)). Understanding the extent of admixture in this region is important for placing admixture in Amboseli in context. In addition, understanding the extent of admixture outside the known hybrid zone is crucial for defining reference panels when calling local ancestry.

 To investigate the structure of our sample, we first used principal components analysis (PCA) to visualize the major axes of variation among high-coverage genomes from Amboseli (n=9), the Aberdares region (n=2), Mikumi National Park in Tanzania (n=11), the SNPRC founders (n=31), and 2 additional SNPRC anubis baboons from the Baboon Genome Project (Fig. 1B) (*14, 23, 24*). We also performed a second PCA on the same set of individuals, plus animals sequenced at low coverage from Mikumi (n=4), the WNPRC (n=6), and Maasai Mara National Reserve in Kenya (n=7) (*15*) (Fig. S4). As expected, in both cases PC1 captures the majority of the variance in the genotype data (84% and 53% proportion variance explained in high and all-coverage genomes, respectively) and separates yellow and anubis baboons, with recently admixed individuals lying between the two extremes. When including a mix of high and low coverage data, samples also separate based on sequencing depth on PC2 and, to a lesser

extent, on PC1 (Fig. S4). Interestingly, however, yellow baboon founders from the SNPRC cluster closely with Amboseli baboons affected by historical admixture (Fig. 1B).

The results of the PCA suggest that founders of the SNPRC colony that were labeled as yellow may contain nonnegligible amounts of ancestry from anubis baboons; indeed, these animals were trapped relatively close to the known hybrid zone, near Kibwezi and Darajani (Fig. 1A). We therefore tested whether we could detect evidence of anubis ancestry by identifying regions of the genome that are identical by descent (IBD) between the SNPRC yellow baboon founders and anubis baboons. To do so, we used IBDmix, a method designed to identify segments in a sample of test individuals (e.g., modern humans, in the original application) that are IBD with a candidate source genome (e.g., Neanderthal in the original application) (*26*). Importantly, this method avoids assuming that members of a given population are unadmixed *a priori* (as has commonly been assumed, for example, for Neanderthal ancestry in human populations of African ancestry). In addition to assessing IBD with anubis baboons among SNPRC yellow baboon founders (n=7), we ran IBDmix on SNPRC anubis baboon founders (n=24) and 2 additional anubis SNPRC colony members, yellow baboons from Mikumi (n=11), and anubis baboons from the Aberdares (n=2).

To assess IBDmix's false positive rates due to either misassignment or incomplete lineage sorting, we also evaluated IBD of yellow baboons with hamadryas baboons (n=2) and Guinea baboons (n=2), which, like anubis baboons, are "northern clade" lineages that separated from the ancestors of yellow baboons ~1.4 million years ago (*14*). There are no known records of wild yellow-Guinea or yellow-hamadryas hybridization in the literature, and the current geographic ranges of these species make gene flow implausible (Fig. 1A). Similarly, we used data from Kinda baboons (n=3) and chacma baboons (n=2)—two "southern clade" baboon lineages that, like yellow baboons, are descended from a lineage that split from the ancestors of anubis baboons ~1.4 million years ago (*14*)—to assess misassignment of yellow baboon IBD segments in anubis baboon populations.

We focused on high quality genotype data for samples analyzed with IBDmix. Specifically, we used the same filtering strategy as in Section 8 (F_4 statistics), except applied to the baboon genome, which retained 28.2 million variants. We then ran IBDmix with default settings, except that we decreased the source individual genotyping error rate to match IBDmix's settings for modern samples (the default settings assume that the source is an archaic genome, whereas all of our samples come from high-coverage, modern samples). We repeated this procedure for each test sample, across all possible source samples, to produce a set of regions that were identified as IBD for each test sample compared to each source sample. We focused on IBD regions that were at least 50 kb long and supported by a LOD score greater than 10 (stringent criteria that are likely to strongly enrich for true shared ancestry (*26*)). We then calculated the mean proportion of each test sample that was inferred to be IBD across possible source individuals. Notably, IBDmix does not differentiate between heterozygous states in which one allele is IBD with the source sample versus homozygous states in which both alleles are IBD. Thus, the proportion IBD reflects the proportion of the genome in which one *or* both alleles (but most likely one) is IBD between test and source individuals.

In yellow baboons from Mikumi, which is far from the known hybrid zone (Fig. 1A), a mean of 5.7 ± 0.8% (s.d.) was estimated to be IBD with anubis baboons, depending on the test and source individuals. This value is likely an overestimate, reflecting incomplete lineage sorting and erroneous calls as well as true shared ancestry. Indeed, 0.9 ± 0.1% of genomes from the same individuals were estimated to be IBD with Guinea baboons and 3.2 ± 0.3% were estimated to be IBD with hamadryas baboons. Given that the range of Guinea baboons is over 5,000 km from Mikumi, we view the most common explanation for IBD calls to be error or incomplete lineage sorting, although historical gene flow between anubis and yellow baboons may partially contribute. A similar pattern was observed in anubis baboons from the SNPRC and one of the two anubis baboons from the Aberdares: low levels of IBD with yellow baboons were estimated genome-wide (5.0 ± 0.8%), but the ranges of these values did not greatly exceed estimates for IBD with other southern clade baboons (geographically distant chacma and Kinda baboons) (Table S8). The second Aberdares anubis baboon was previously identified as an admixed individual by the Baboon Genome Consortium, which reported ~546 Mb of yellow baboon-derived nuclear DNA (*14*). Our analysis supports that finding, with an IBD estimate for one or both alleles for 15.2% of the genome.

In contrast to the other non-Amboseli populations, yellow baboon founders from the SNPRC colony clearly appear to be admixed. IBDmix estimates an average IBD of 26.5 ± 6.7% for SNPRC yellow founders with anubis source individuals, in contrast to 5.7 ± 0.8% for Mikumi yellow baboons (t-test p = 1.68 x 10^{-4}). Estimated IBD for SNPRC yellow founders with anubis baboons also exceeds the estimates for SNPRC yellow IBD with Guinea baboons or hamadryas baboons (vs. Guinea baboons: mean difference = 20%, paired t-test p = 3.37 x 10^{-5}; vs. hamadryas baboons: mean difference 12%, p = 7.16 x 10^{-5}). Further, when regions of the genome in SNPRC yellow founders were inferred to be IBD against one anubis source individual, this inference tended to be highly consistent across other anubis source individuals, leading to extended "peaks" where IBD is broadly supported across most or all source animals. Because IBDmix is agnostic to the direction of gene flow, we therefore also observe a large fraction of anubis baboon genomes that appear to be IBD with yellow baboons when SNPRC yellow founders are used as source individuals (17.6-30.8%; Table S9). This value drops to 3.1-6.9% if Mikumi yellow baboons are used as source individuals.

9.2. Identifying introgressed ancestry using LCLAE confirms admixed status of SNPRC yellow founders

We complemented our IBDmix analyses by testing all yellow and anubis baboon samples for admixture using LCLAE, which (unlike IBDmix) distinguishes between heterozygous and homozygous ancestry states and allowed us to include low coverage samples. For each individual, we called local ancestry states using reference panels for anubis and yellow baboon allele frequencies composed of all yellow and anubis individuals except the focal individual. LCLAE assigned all high coverage samples from the SNPRC anubis founders and Mikumi yellows, and one of the anubis baboons sampled from the Aberdares, to their *a priori*-defined ancestries for >95% of the genome. Low coverage samples were similarly categorized, but low coverage appears to pull admixture proportions away from the extremes (i.e., in Mikumi, the mean proportion of anubis

ancestry is $3.3 \pm 0.2\%$ (s.d.) in high coverage samples, but $14.3 \pm 6.4\%$ in low coverage samples: see Section 7.1 for direct estimates of coverage-induced error based on these samples). As expected based on IBDmix and PCA results, high coverage SNPRC yellow founders were estimated to have $17.1 \pm 4.2\%$ anubis ancestry, and the high coverage admixed anubis animal from the Aberdares was estimated to have 11.0% yellow baboon ancestry (Fig. S5A).

For comparison between LCLAE and IBDmix, we collapsed homozygous and heterozygous introgressed ancestry estimates from LCLAE because IBDmix does not differentiate between these two alternatives. Genome-wide estimates of introgressed (i.e., IBD) ancestry are highly correlated between LCLAE estimates and IBDmix estimates (Pearson's $r > 0.99$, p-value $< 10^{-21}$ for Mikumi and SNPRC yellow baboons; $r > 0.99$, p-value $< 10^{-20}$ for Aberdares and SNPRC anubis founders; Fig. S5B-C). IBDmix tends to detect more shared ancestry than LCLAE, although this effect lessens if IBD is only inferred for loci where the results for most (at least 70%) source individuals agree. On a site-specific level, regions identified using LCLAE were more than 20x as likely to be IBD based on IBDmix analysis across species, relative to the genomic background (i.e., randomly sampled, size-matched regions of the genome).

9.3. Introgression versus incomplete lineage sorting

Four lines of evidence suggest that the patterns observed in Amboseli and the SNPRC yellow founders primarily reflect introgression rather than incomplete lineage sorting (ILS). First, in recent hybrids, the inference of introgression based on admixture is directly supported by field observations in Amboseli and elsewhere (*22, 54*). Second, only specific populations, near the documented hybrid zone, exhibit high levels of introgressed DNA; because ILS should affect all populations equally, this pattern points to population-specific histories of gene flow. Specifically, in Amboseli and in the SNPRC yellow founders, estimates of IBD (based on IBDmix) or anubis ancestry (based on LCLAE) are ~4-5-fold higher than for yellow baboons from Mikumi (Section 9.2). Third, ILS predicts similar levels of genome sharing between yellow baboons and all three northern lineage baboon species (Guinea and hamadryas), as their common ancestors diverged from the ancestors of yellow baboons on a single branch of the phylogeny. However, inferred IBD for SNPRC yellow founders and Amboseli yellow baboons is 12% and 20% higher with anubis baboons than with hamadryas and Guinea baboons, respectively (Section 9.1). This observation contrasts with results for the Mikumi yellow baboons, where IBD with anubis baboons is only 0.9% and 2.5% greater than for hamadryas and Guinea baboons, respectively. Fourth, our estimates of admixture proportions using the F_4-ratio draw on methods that were specifically developed to differentiate between admixture and ILS (*79*) and produce highly similar estimates to LCLAE-based estimates of overall admixture in Amboseli samples (within 1.75% difference; Section 8).

Finally, we note that our local ancestry calls are unlikely to be influenced by ILS due to the approach for ancestry tract calling implemented in LCLAE. LCLAE calls for each genomic window can be interpreted as the relative likelihood that the test individual's genotype is drawn from source population *A*, source population *B*, or a combination of both (heterozygous ancestry). ILS may increase the difficulty of ancestry

assignment in a given region (by reducing the divergence between yellow and anubis baboons), but it is unlikely to introduce systemic bias.

10. Timing of admixture

To estimate the timing of admixture in high coverage samples from Amboseli, we used DATES (*35*). This method estimates the timing of admixture by fitting an exponential regression to the decay in ancestry covariance by genetic distance. While DATES was initially developed to date ancient DNA samples, the model and interpretation of DATES is no different when using modern samples (Priya Moorjani, personal communication). Thus, the output from DATES can be interpreted as "generations before the age of this sample", where sample age is 0 for modern data. We evaluated admixture timing in seven "historic" Amboseli hybrids and two "recent" Amboseli hybrids sequenced at medium to high coverage. In all cases, we considered a simple model of a single pulse of admixture. Although we believe that multiple waves of admixture have likely influenced baboon evolution in East Africa, current methods are not able to differentiate separate immigration events from the same source population, especially when the level of discontinuity in these immigration waves is unclear. Thus, we view the DATES estimates as composite estimates of the timing of anubis-yellow admixture across the genome, weighted towards more recent events.

To run DATES, we used the SNPRC anubis founders and Mikumi yellow baboons to represent the source of anubis and yellow ancestry for Amboseli baboons. We did not use SNPRC yellow founders for this analysis because DATES requires contiguous genome sequence to run; continuous yellow ancestry in the SNPRC yellows is interrupted by introgressed DNA. After filtering for common variants called in high coverage Amboseli samples (minor allele frequency ≥ 5%), we converted SNP locations into a genetic map using baboon recombination rates (*24*) (Section 12.3). Specifically, we calculated the mean recombination rate per 100 kb, removed unreasonably large local recombination rates (>100x larger than the median recombination rate), and then scaled LDhelmet's population-scaled recombination rate to centiMorgans based on the baboon chromosome lengths estimated by Cox et al. (*88*). We then fit a loess smoothing function to predict the position (in Morgans) of each SNP along the genome and used DATES to estimate the timing of admixture, using a maximum distance of 0.5 Morgans and a bin size of 0.001 Morgans.

Consistent with long-term field observations in Amboseli, the number of generations since admixture estimated for recent hybrids (5 and 21 generations) is substantially smaller than for historic hybrids (Fig. S6A). This inference is also consistent with longer heterozygous and homozygous anubis ancestry tracts in individuals with known recent anubis ancestors, compared to Amboseli animals with no known recent anubis ancestors (Fig. 3B; Fig. S6B).

11. Verifying historic cessation of gene flow using pseudo-hybrid PSMC

DATES estimates admixture timing regardless of the actual model for gene flow. Previous work has inferred a complex history of contact, probably involving repeated bouts of gene flow, rather than a simple isolation-migration model or continuous gene flow (*15*). Alternatively, the capacity for anubis and yellow baboons to successfully interbreed could be consistent with continuous exchange between species, at least where

their ranges meet, and call into question our treatment of anubis and yellow baboons as distinct taxa. To verify cessation of gene flow in the history of yellow and anubis baboons, we therefore used pseudo-hybrid pairwise sequential Markovian coalescent (PSMC) modeling (*89*). Pseudo-hybrid PSMC uses artificial genomes where half of the genome is constructed from one parent taxon and the other half is constructed from the other parent taxon. Truly divergent taxa produce PSMC plots that show a sudden increase in effective population size when parent lineages begin evolving independently.

Following (*89*), we created haploid files for each high coverage SNPRC anubis baboon founder (n=24) and Mikumi yellow baboon (n=11) by picking one base for each position using samtools mpileup (*81*) and *pu2fa* (*90*). These populations were selected because of the availability of high coverage genomes and the absence of evidence for recent gene flow. We used *psmcfa_from_2_fastas.py* (*89*) to create a merged psmcfa file for each pair of yellow and anubis baboons (264 unique combinations), and then ran *psmc* with default parameters (*91*). Results were collected across pairs of individuals using *PSMC_emit_last_iteration_coord.py* (*89*) assuming a generation time of 11 years and a mutation rate of 0.57×10^{-9} mutations per base pair (directly estimated for baboons in (*92*)).

As expected, we observed a near-asymptotic increase in estimated effective population size (Fig. S7), consistent with historical cessation of gene flow between species. These results produce an estimate for the minimum effective population size (i.e., cessation of gene flow) around 250 – 300k years ago. However, these dates are unreliable, as translating between pseudo-hybrid PSMC plots and actual dates assumes the same effective population size for both descendant taxa, which is violated for yellow and anubis baboons based on recent estimates (see (*14*)). Nevertheless, the pseudo-hybrid PSMC results support divergence between anubis and yellow baboons minimally hundreds of thousands of years ago, even if contact and reconvergence occurred after the initial divergence between northern and southern baboon lineages.

12. Evidence for selection against anubis introgression

To investigate whether signatures of selection against archaic introgression in humans are paralleled in baboons, we performed three tests. First, we asked whether regions of the genome that exhibit greater divergence between the parent species have reduced introgressed (anubis) ancestry in Amboseli. In humans, loci that contain fixed or near-fixed differences between humans and Neanderthals are more likely to be depleted of archaic ancestry (*27*). Second, we tested whether regions of the genome that are predicted to be more affected by background selection also exhibit reduced introgressed ancestry, as reported for Neanderthal introgression into the human genome by Sankararaman et al. (*28*). Third, we tested whether introgressed ancestry is depleted in regions of the genome with low mean recombination rate, as expected if deleterious alleles are selectively eliminated along with linked sequence. This pattern is also observed for Neanderthal and Denisovan ancestry in modern humans as well as in other hybridizing animal and plant taxa (*29, 31, 32, 93, 94*). For all three analyses, we calculated mean anubis ancestry for non-overlapping windows of the genome in (i) all baboons from Amboseli (n=442); (ii) historical hybrids only (n=214); and (iii) recent hybrids only (n=188) (see Section 1.2 for how recent and historical hybrids were defined). We report results from local ancestry calling with LCLAE in the main text.

However, the results are qualitatively unchanged if we use ancestry assignments based on Ancestry HMM (*74*) (Table S3). In the main text, we report results using window sizes of 250 kb, but our results are also robust when using window sizes from 100 kb up to 1 Mb (Table S3).

12.1. Introgressed ancestry as a function of divergence between parental taxa

Using all non-Amboseli yellow and anubis baboons (after masking loci potentially affected by introgression: Section 6.2), we counted the number of fixed differences that occurred in non-overlapping 250 kb windows across the genome. As reported in the main text ("Selection against introgression in Amboseli"; Fig. 2), we observed a strong negative relationship between introgressed ancestry in these windows and the number of fixed differences between yellow and anubis baboons (Spearman's rho = -0.119, p = 6.44 x 10^{-31}), which is largely driven by historical hybrids (Fig. 2A, 3D). As an alternative approach, we also identified highly differentiated sites based on allele frequencies estimated from anubis baboon and yellow baboon reference panel individuals (Section 6.2). We calculated F_{ST} for every biallelic variant called in at least 5 anubis and 5 yellow baboons (n=21 million biallelic SNPs) as the loss in heterozygosity between species relative to a panmictic population (*95*). Specifically, $F_{ST} = \frac{Ehet_{panmictic} - mean(Ehet_{populations})}{Ehet_{panmictic}}$, where *Ehet* (the expected heterozygosity) is calculated for each population and for a hypothetical panmictic population from observed allele frequencies, assuming Hardy-Weinberg equilibria. Our estimates of F_{ST} are broadly consistent with previous estimates based on genome-wide biallelic SNP data (*15, 24*), although they are larger than previously reported estimates based on a limited set of ~12 multiallelic microsatellite markers originally typed for paternity analysis (*22*). Introgressed ancestry is also negatively correlated with the number of highly differentiated sites (Fig. S8). Unsurprisingly, this relationship is strongest for nearly fixed differences ($F_{ST} \geq 0.75$, Spearman's rho = -0.109, p = 1.21 x 10^{-28}) and weaker, although still significant, for sites that are less differentiated (e.g., $F_{ST} \geq 0.5$; Spearman's rho = -0.042, p = 2.12 x 10^{-5}).

Two additional complications are relevant here. First, if yellow and anubis populations in our sample have also been affected by past admixture (beyond what we were able to detect and mask) and if the regions recalcitrant to introgression in modern baboons were also resistant to introgression in the past, then divergence estimates based on these other populations would be correlated with reduced introgression in Amboseli. However, such a scenario would still be consistent with selection against introgression. Second, divergence could affect our ability to estimate local ancestry in Amboseli. If so, we would expect greater power to identify introgressed anubis segments in Amboseli for more diverged regions of the genome. This would result in higher levels of anubis ancestry in Amboseli for more diverged regions, the opposite pattern to what we observed.

12.2. Introgressed ancestry as a function of estimated background selection

To estimate the degree of background selection across the *Panubis1.0* baboon genome (*65*), we used the B value statistic from McVicker et al. (*30*). B values are estimates of the proportion of nucleotide diversity maintained in the face of purifying selection, as a function of the deleterious mutation rate, the strength of selection, and

recombination rate. These values assume that selection acts multiplicatively over loci and that the strength of selection is sufficient to ignore deleterious alleles in homozygous states (*30*). B values near 0 represent strong background selection (i.e., near complete removal of diversity due to natural selection at linked sites), whereas large values correspond to little background selection.

Following McVicker et al. (*30*) and Sankararaman et al. (*28*), we distinguished two classes of sites under selection: base pairs in exons (including UTRs) and base pairs in introns or promoter regions (defined as 10 kb upstream of gene transcription start sites). Exonic regions were assigned a deleterious mutation rate of 7.4×10^{-8} mutations per base pair, and introns and promoters were assigned a deleterious mutation rate of 8.4×10^{-10} mutations per base pair. Selection coefficients were drawn from an exponential distribution with mean 2.5×10^{-3} and mean 1.0×10^{-5} for exons and introns/promoters, respectively. These values are based on the human genome to maximize comparability, but we also considered a range of selection coefficients from 1×10^{-5} to 5×10^{-2}, which generated the same qualitative results. We note that, although our estimates are based on the anubis baboon reference genome, there is a high degree of synteny documented for cercopithecine primates (e.g., baboons and gelada monkeys (*96*), baboons and rhesus macaques (*97*)).

As reported in the main text ("Selection against introgression in Amboseli"; Fig. 2), we observe reduced introgressed ancestry in 250 kb windows of the genome with higher levels of linked selection (lower B values). This result was largely unaffected by the choice of selection coefficients (all rho > 0.16, $p < 5 \times 10^{-64}$), likely because our analyses are focused on relative rather than absolute values of B. This result is stronger in historical hybrids than for those with recent anubis ancestry, and robust to changes in window size (Table S3). Regions of the genome that contain more variable sites (suggesting weaker background selection) also have higher levels of introgressed ancestry (Spearman's rho = 0.196, $p = 4.83 \times 10^{-90}$ for 250 kb windows).

12.3. Introgressed ancestry as a function of local recombination rates
To evaluate the relationship between introgressed ancestry and local recombination rates, we required a fine-scale recombination map for baboons. An LD-based recombination map is available for the anubis baboon from Robinson et al. (*24*), but the published version is linked to an older version of the baboon genome (*Panu2.0* (*14*)) than the version we used here (*Panubis1.0* (*65*)). We therefore regenerated the anubis baboon recombination map using *LDhelmet* (*98*) (v1.9) and resequencing data from the 24 SNPRC anubis baboon founders (*24*). We first phased the genomes using *Beagle* (*99*) (v5.0), after excluding uninformative singletons and pruning variants to at least 10 bp distance apart. We then ran *LDhelmet* with a block size penalty of 5 (as recommended by Chan et al. (*98*)), an $N_e = 40,000$ (based on (*100*)), and default parameter values except where noted. We set the population-scaled mutation rate, θ, to 0.0016 in baboons, which is an approximation of the value of θ calculated from the number of variants in the input data prior to pruning. We used a window size of 50 and ran the Markov chain Monte Carlo inference for 1×10^6 iterations, following a burn-in period of 10^5 iterations. The resulting estimates are in terms of ρ (= $4N_e r$), the population-scaled recombination rate. We converted these values to centiMorgans based on the lengths of baboon chromosomes reported by Cox et al. (*88*).

As reported in the main text ("Selection against introgression in Amboseli"; Fig. 2), we observed increased introgressed ancestry in 250 kb windows of the genome with increased mean recombination rates. This result is also driven by historical hybrids rather than individuals with recent anubis ancestry, and robust to changes in window size ranging from 100-1000 kb (Table S3). Although our recombination map is derived from anubis baboons, our results are consistent using large window sizes (Table S3), where the recombination landscape is likely to be well-conserved between species (*101*).

12.4. Depletion of anubis ancestry in putatively functional regions of the genome

We tested whether anubis ancestry was depleted in regions of the genome with known functional importance. We focused on coding sequences, gene promoters, and enhancers, in each case comparing anubis ancestry levels to size-matched, putatively neutral regions more than 100 kb away from annotated elements (i.e., genes, promoters, and enhancers). We used gene annotations from NCBI to define coding sequences (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/008/728/515/GCF_008728515.1_Panubis1.0/GCF_008728515.1_Panubis1.0_genomic.gtf.gz) (*102*), and defined promoters as the 10 kb upstream of the 5'-most transcription start site. Because baboon enhancer annotations are not available, we defined putative baboon enhancers by projecting coordinates from ENCODE H3K4me1 ChIP-seq data for human peripheral blood mononuclear cells (*103*) onto the *Panubis1.0* genome using the UCSC Genome Browser *liftover* tool (*104*) and a liftover file we generated (available on Zenodo (*46*)). We then identified five sets of size-matched regions using the shuffle command in *bedtools* (*105*) (version 2.25.0), excluding regions that were within 100 kb of annotated genes, promoters, or enhancers. For each type of element and each individual, we calculated the proportion of anubis ancestry and estimated mean anubis ancestry across all elements, yielding one estimate per individual. For each element type, we then compared these values to size-matched, putatively neutral regions using paired t-tests (n=442 individuals).

13. Comparative evidence from archaic hominin admixture

To provide context for our findings, we compared them to evidence for selection against archaic introgression in humans, focusing specifically on admixture with Neanderthals. All three of the tests we used have been previously reported in the literature (*27-29*), but using different human reference genomes and differing amounts of information (based on what was available at the time) on Neanderthal-human divergence and introgression. We therefore re-performed them here, which allowed us to follow an analysis pipeline that was as parallel to our analysis of the baboons as possible.

To do so, we first downloaded Neanderthal ancestry calls for HapMap CEU individuals from (*39*). Following (*29, 39*), we converted posterior probabilities for Neanderthal ancestry to 0/1 ancestry calls based on a threshold of 0.42 for Neanderthal ancestry (this approach assumes that homozygous Neanderthal tracts, i.e., ancestry state 2, are rare). We then averaged ancestry states over all sites in a given genomic window for each individual and averaged these per-individual values to obtain an estimate of introgressed (Neanderthal) ancestry in the 1000 Genomes CEU population (*106*). Importantly, population-level introgressed archaic ancestry across the genome is highly

correlated with the haplotype-based calls of Sankararaman et al. (*28*) (at the 50 kb scale, Spearman's $\rho = 0.88$; at the 500 kb scale, $\rho = 0.94$).

B values for the human genome were obtained from (*30*) and converted to hg19/gr37 coordinates using the Ensembl AssemblyConverter (*107*). To calculate window-based recombination rates, we used the HapMap Project recombination maps for CEU individuals (*108*). Finally, to identify fixed differences between modern humans and Neanderthals, we downloaded and merged genotype calls for 3 high coverage Neanderthals (the Altai Neanderthal (*109*), the Vindija Neanderthal (*110*), and the Chagyrskaya Neanderthal (*111*)) from (*112*). We then identified fixed differences between Neanderthals and modern humans by filtering for single nucleotide variants that were called in at least 2 of the high coverage Neanderthal samples and fixed for the alternate allele in Neanderthals but the reference allele in modern humans (excluding African ancestry populations). We removed variants that were variable in African populations in the 1000 Genomes Project (*106*). This left 769,488 variants that appear to be fixed differences between Neanderthal and non-African modern human genomes, and are at the very least highly differentiated between non-African modern humans and Neanderthals.

We calculated the proportion of Neanderthal ancestry for the CEU population for non-overlapping 250 kb windows across the genome. We excluded windows in which no alleles were called for at least two of the three Neanderthal genomes. These proportions are negatively correlated with the number of fixed differences, positively correlated with B values (i.e., reduced background selection), and positively correlated with local recombination rate, as expected based on previous reports (*27-29*) (Fig. 2; Table S2).

14. Ancestry-associated gene expression levels

14.1. Data processing
To investigate the contribution of genetic ancestry to gene expression, as well as the relationship between ancestry effects and signatures of selection, we used RNA-seq gene expression data from 157 samples, representing 145 unique individuals. All samples were collected from reproductively mature adults, where we were able to control for age, sex, and batch/sampling effects in the analysis. Further, nearly all animals were sampled in consistent ecological conditions (i.e., during the long dry season in Amboseli, when darting is most feasible). Data from 63 samples were obtained from whole blood and reported in (*40*); data from 94 samples were obtained from white blood cells and reported in (*41, 42*) (Table S1). All data were processed together following a common pipeline. Specifically, data were trimmed using TrimGalore (*63*) with default parameters while retaining reads with trimmed length > 25 bases. We then mapped reads from each sample to *Panubis1.0* (*65*) using STAR 2-pass mapping and a combined splice junction database created from all data during the first round of mapping (*113*). Gene-level counts were estimated using HTSeq (*114*).

Whole blood samples from (*40*) were sampled, extracted, and sequenced following a different protocol for white blood cell samples collected after 2013. We therefore analyzed the whole blood (n=63) and the white blood cell data sets (n=94) separately. In each case, we filtered for genes with mean TPM ≥ 2 that had non-zero counts in at least half of the samples in the data set, resulting in a set of 11,050 and 11,238 expressed genes

in the whole blood and white blood cell data sets, respectively. We then removed genes where local ancestry was nearly fixed (mean anubis ancestry < 10% or > 90%), resulting in 10,615 (whole blood) and 10,798 (white blood cell) genes analyzed for local ancestry effects on gene expression. To assign gene-level local ancestry values to each individual, we calculated the mean number of anubis alleles for a window spanning the gene transcription start site to transcription end site, plus an additional flanking 10 kb on each side. For very short genes, we extended the total window size to 50 kb, as local ancestry calls are less reliable for short windows.

Finally, we normalized the gene-level counts matrix for each data set separately using *voom* (*115*), controlled for library size with *edgeR* (*116, 117*) (version 3.30.3; function *calcNormFactors*), and modeled the normalized gene expression data in *limma* (*118*) (version 3.44.3; function *lmFit*) as a function of known batch effects (year of sampling and, for white blood cells only, cell type composition summarized using the first two principal components of flow cytometry-based abundance data for $CD3^+CD4^+$ T cells, $CD3^+CD8^+$ T cells, $CD3^-CD20^+$ B cells, $CD3^-CD14^+$ monocytes, and $CD3^-CD16^+$ natural killer cells, as described in Lea et al. (*41*) and Anderson et al. (*42*)). We then used the residuals of this model to investigate genetic ancestry effects on gene expression.

14.2. *Testing for additive effects of local genetic ancestry on gene expression levels*

We modeled the normalized, batch-corrected gene expression data using the linear mixed model approach implemented in *emmreml* (*119*) (version 3.1). We fit fixed effects of local genetic ancestry around each gene, global genetic ancestry, age, and sex, as well as a random effect to control for relatedness between samples. Relatedness estimates (i.e., the K matrix in the mixed model) were calculated from RNA-seq genotype calls (see Anderson et al. (*42*)) as the covariance of the 0/1/2 genotype matrix. To control for multiple hypothesis testing across genes, we calculated the false discovery rate following (*120*), against an empirical null distribution derived from 50 permutations of each variable of interest (either global ancestry or local ancestry) (*121*). Following the discovery that genome-wide ancestry was not associated with variation in gene expression levels, we further tested whether the effects of genome-wide ancestry were subsumed within the random effect component of the model. However, when we repeated the analysis using a K matrix defined only by pedigree data, which captures familial relatedness but not ancestry, we still did not detect global ancestry effects on gene expression.

14.3. *Testing for non-additive effects of local genetic ancestry on gene expression levels*

Because hybridization has been reported to generate transgressive gene expression patterns in other species pairs (i.e., expression levels outside the range of the parent taxa (*122-126*)), we also investigated possible non-additive associations between local ancestry and gene expression levels. To do so, we subtracted out the random effect of kinship using *emmreml* before using the *segmented* package (*127*) (version 1.2-0) to fit a model with different slopes for anubis-like and yellow-like baboons compared to heterozygous individuals. Overall, we found very little evidence for non-additive effects in our sample. Only five genes were both significantly associated with local ancestry and had different slopes for ancestry states 0 to 1 versus 1 to 2 (10% FDR), in either the

whole blood or white blood cell data sets. Thus, while local ancestry effects on gene expression in Amboseli are common, most appear to act additively.

We note that our finding of rare non-additive ancestry effects on gene expression may be because anubis baboons and yellow baboons are still relatively closely related (~1.4 million years diverged (*14*)) compared to the cases reported in the literature (*122-126*). Thus, they still share substantial genetic variation and form viable and fertile hybrids. For example, recently diverged zebra finch subspecies (which also hybridize without obvious phenotypic costs) also exhibit gene regulatory divergence but show no evidence of dysregulation in first-generation hybrids (*128*). However, it may also be the case that we missed cases of non-additivity that would have been detected in first generation hybrids, as recombination can break apart incompatible combinations in later generations (*129*).

14.4. *Signatures of selection are stronger near genes with regulatory divergence*

Despite greater genetic divergence between Neanderthals and modern humans, introgressed Neanderthal variants found in modern humans exhibit similar effects on gene expression levels as variants that arose in the modern human lineage (*37*) (although high frequency Neanderthal alleles may be more likely to affect expression (*130*)). McCoy and colleagues (*37*) proposed that the Neanderthal-derived variants that persist in modern humans today represent a non-random subset of those first introduced by admixture, following selection that removed introgressed variants with large effects on gene regulation. However, as only partial Neanderthal genomes, and only a subset of Neanderthal alleles, are present in modern humans (*39, 131*), it is no longer possible to directly test this hypothesis in humans (but see Colbran et al. (*38*) for predictions in introgression deserts based on variation segregating in modern humans).

It is possible, however, to directly test whether this hypothesis holds for admixture in baboons. If it does, we predict that genes with greater functional divergence (i.e., where local ancestry has large effects on gene expression) will have lower levels of introgressed anubis ancestry than genes that are not functionally diverged. To test this prediction, we rank-ordered all genes in our gene expression analysis by the magnitude of their estimated local ancestry effect (irrespective of sign). Because we generated two estimates for most genes (one for the whole blood RNA-seq data set and one for the white blood cell RNA-seq data set), we first used multivariate adaptive shrinkage (*mashr* (*132*)) to refine our effect size estimates before calculating the mean effect size across the two data sets. We then calculated the mean ancestry per individual for quantiles of genes with the largest and smallest effect of local ancestry on gene expression (across quantiles from 10-40%, in intervals of 5%; Table S4). We tested whether mean anubis ancestry was different between the two sets, using a paired t-test where mean values for the largest/smallest effect size genes were represented for each individual. As reported in the main text and Table S4, we find consistent evidence that functionally diverged genes contained less introgressed ancestry than genes that are not functionally diverged.

We also predict that signatures of selection will be strongest near genes with greater functional divergence. To test this prediction, we used the magnitude of the correlation between recombination rate and introgressed ancestry as a proxy for the strength of selection (as in Fig. 2E, Fig. 3F, "Selection against regulatory divergence"). We first replicated our earlier whole-genome analyses (Section 12.3) for genic regions

(transcription start site to transcription end site, plus an additional flanking 10 kb on each side, and extended to a minimum total window size of 50 kb for small genes). We confirmed that higher local recombination rates also predict higher levels of anubis ancestry in and around protein-coding genes (n=19,050 genes; Spearman's rho = 0.116, p = 1.13 x 10^{-58}), with an effect size similar to that estimated for the whole genome (rho = 0.127). This pattern is also detectable among the subset of genes that we tested for ancestry-associated expression (n=10,168; rho = 0.111, p = 1.29 x 10^{-29}). In both cases, signatures of selection were again stronger in historically admixed individuals than in recently admixed individuals (all genes: rho = 0.115 vs. rho = -0.013; genes tested for local ancestry-associated gene expression: rho = 0.110 vs. rho = -0.029).

To test whether genes with larger effects of local genetic ancestry on gene expression are associated with stronger signatures of selection against admixture, we calculated the correlation between local recombination rate and anubis ancestry for quantiles of genes with the largest and smallest associations between genetic ancestry and expression (again across quantiles from 10% to 40%, in intervals of 5%). For each quantile comparison, we used a bootstrapping approach to estimate Spearman's rho for the genes with the largest and smallest effect sizes, as well as the difference in rho between these two gene sets (n=10,000 repetitions). For the more extreme quantiles ($\leq$ 20%), genes with large local ancestry effect sizes consistently exhibited stronger recombination rate-introgressed ancestry correlations than genes with near-zero effect sizes (Fig. 4B; Table S4).

15. Predicting anubis ancestry across the genome and through time

15.1. Genomic features predict introgressed ancestry

To evaluate the capacity to predict the landscape of admixture from genomic features alone, we used *glmnet* (*133*) (version 4.0-2) to predict population-level introgressed ancestry in 250 kb non-overlapping windows of the baboon genome (n=10,324 windows, after excluding the 50 kb at the chromosome ends and windows where the recombination rate was estimated to be greater than 100x the median recombination rate). Because the predictive accuracy of our models was insensitive to the elastic net mixing parameter, we used a mixing parameter of 0.5 (in *glmnet*, 0 corresponds to ridge regression and 1 corresponds to LASSO (*134*)). For 200 iterations of *glmnet*, we fit the elastic net model to a random 75% of genomic windows and used the resulting parameter estimates to predict introgressed ancestry in the remaining 25% of windows. We repeated this analysis after permuting introgressed ancestry values to establish a null expectation, and for recent and historical hybrids separately to investigate our relative predictive accuracy.

We included 32 features in our models (Table S5). These features included (i) the number of common genetic variants (minor allele frequency $\geq$ 5%, based on the full sample of non-Amboseli yellow and anubis baboons) and highly differentiated sites (F_{ST} > 0.5-1.0, as described in Section 12.1); (ii) the B value statistic (see Section 12.2); and (iii) the local recombination rate based on the anubis baboon genome (see Section 12.3), incorporated as three correlated features (a raw rate, a ranked value, and a $\log_{10}$-transformed value). We also calculated (iv) the proportion of each window that overlapped gene exons, gene bodies, predicted-PBMC enhancers (see Section 12.4), and

CpG islands, as well as (v) window GC content. CpG islands were annotated using the *emboss* cpgplot function with default parameters to identify windows of the genome longer than 200 bases with greater than 50% GC content and an observed/expected CpG ratio > 0.6 (*135, 136*). GC content was obtained for each 250 kb window using the nuc function in *bedtools* (*105*) (v2.25.0). The mean effect size estimate for each feature can be found in Table S5.

15.2. Genomic features predict the rate of change in anubis allele frequencies

Anubis ancestry has increased in Amboseli since the early 1980's, when anubis and anubis-like immigrants began arriving in the study population (*18*) (Fig. 3A). To evaluate whether changes in anubis ancestry since that time are predictable along the genome, we estimated the anubis ancestry proportion in Amboseli for 100 kb genomic windows (excluding the 50 kb at the chromosome ends; n=25,875 total windows), in each chronological year from 1979-2020. We based this calculation on individuals within the Amboseli resequencing data set who were present in the population in that year, according to near-daily demographic records (n=11-228, mean=135.19 resequenced individuals per year). An individual's ancestry was included in the annual population estimate after entering the study population (through birth, immigration, or the onset of observation). They were removed after disappearance from the study population (due to death, dispersal, or because we stopped monitoring their group). For each window, we fit the annual population anubis ancestry as the outcome variable in a linear model with year (1979 – 2020) as a continuous predictor. We estimated the annual change in anubis ancestry (Δ_{anubis}) as the parameter estimate for the year effect.

We then modeled Δ_{anubis} across windows as a function of the starting anubis ancestry frequency in 1979, mean recombination rate ($\log_{10}$ transformed), weighted F_{ST}, and B values (Table S6). Mean recombination rates were calculated as described in Section 12.3, F_{ST} values were calculated as described in Section 7.3, and B values were calculated as described in Section 12.2. We excluded windows in which recombination rate estimates were greater than 100x larger than the median recombination rate and windows that had no variants (so F_{ST} was not calculated), resulting in a final data set of 25,797 windows. We also confirmed that estimates of model effects were generally stable across window sizes from 35 - 250 kb.
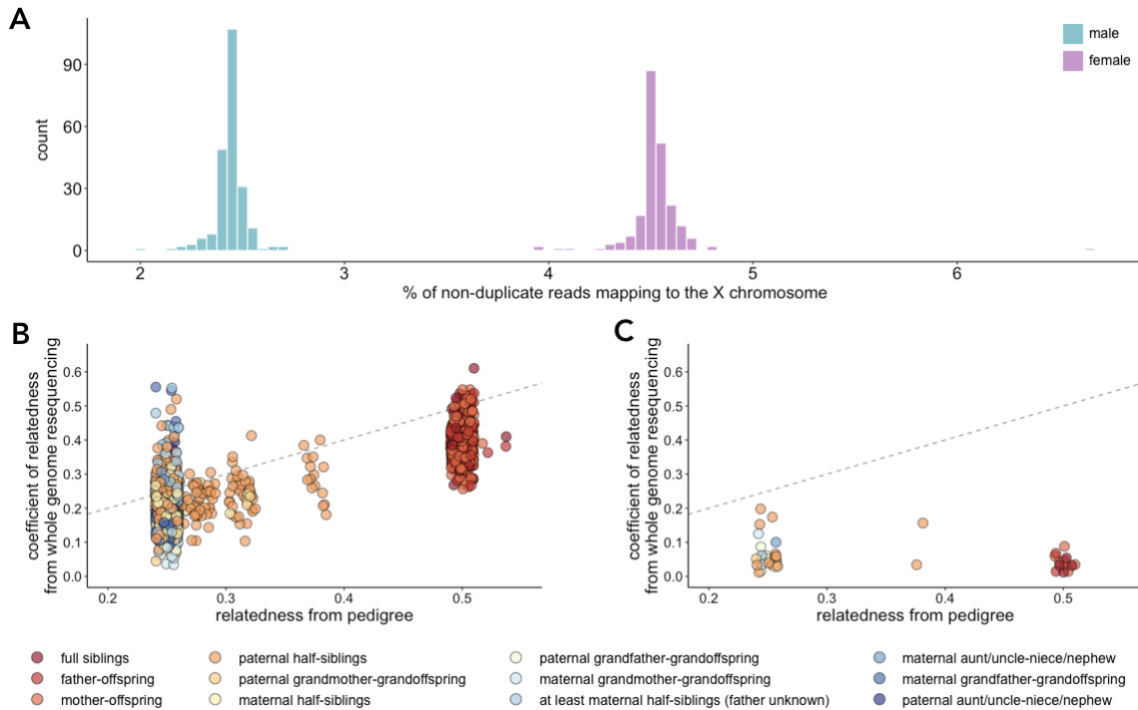
Fig. S1: Quality control for individual identity in the Amboseli resequencing data.

(A) For sex, we confirmed that resequencing data from females (pink) had approximately 2x the percentage of non-duplicate mapped reads assigned to the X chromosome relative to males (blue). **(B-C)** Relationship between relatedness estimates from the Amboseli Baboon Research Project pedigree and from genome resequencing using *lcMLkin* (*68*) for pairs of individuals with expected pedigree-based relatedness ≥ 0.25 (n=382 unique individuals). **B** shows dyads where both individuals were correctly identified (n=376 unique individuals), resulting in a strong correlation between pedigree and sequencing-based relatedness (Pearson's r = 0.789, p < 10^{-300}). **C** shows dyads which include at least one of 6 incorrectly identified individuals, where pedigree and sequencing-based relatedness estimates are not correlated (Pearson's r = -0.244, p = 0.129). Cases of expected pedigree-based relatedness equal to 0.5 (e.g., parent-offspring, full siblings) are colored in dark orange or red. The grey dashed line corresponds to the y = x line. The considerable noise in this relationship likely reflects noise in relatedness estimates from low coverage resequencing data, structure from the admixed nature of the population, and some degree of cryptic relatedness not apparent in the pedigree data.
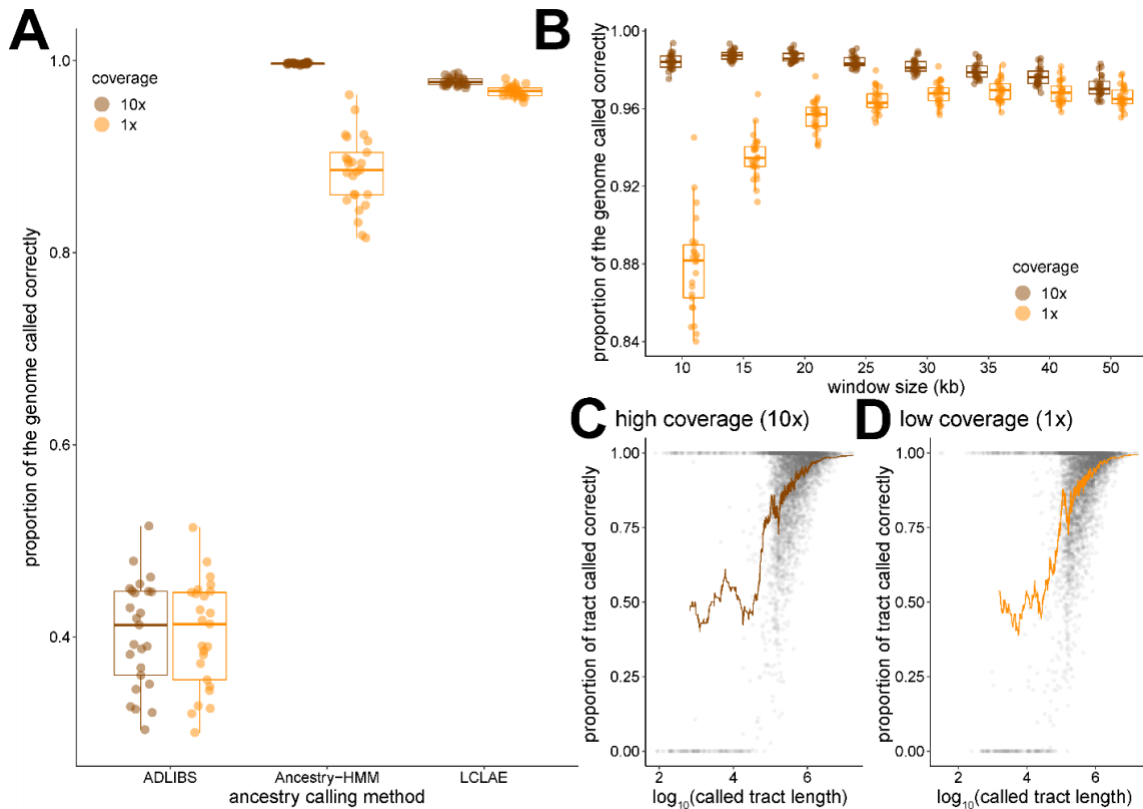
Fig. S2: Accuracy of local ancestry assignment.
(A) Comparison of local ancestry calling using AD-LIBS (*75*), Ancestry HMM (*74*), and LCLAE (*15*) with both high (10x, brown) and low (1x, orange) coverage data for 25 simulated individuals. **(B)** The proportion of the genome called correctly using LCLAE as a function of the window size for calling ancestry and sequencing coverage. The smallest window sizes are less accurate using low coverage data because few ancestry informative sites are available for making assignments, while the largest window sizes are slightly less accurate than the maximum accuracy for both levels of coverage because they miss small ancestry tracts. **(C-D)** The relationship between called tract length and accuracy in calling local ancestry with LCLAE for high (C) and low (D) coverage data. Each point represents an ancestry tract called for a simulated individual, and the brown (C) and gold (D) lines represent the moving mean accuracy for non-overlapping sets of 70 ancestry tracts. Tracts less than 1 kb perform little better than chance. In contrast, longer tracts are close to 95% accurate, with the main source of error due to missed small ancestry tracts that are contained within large tracts.
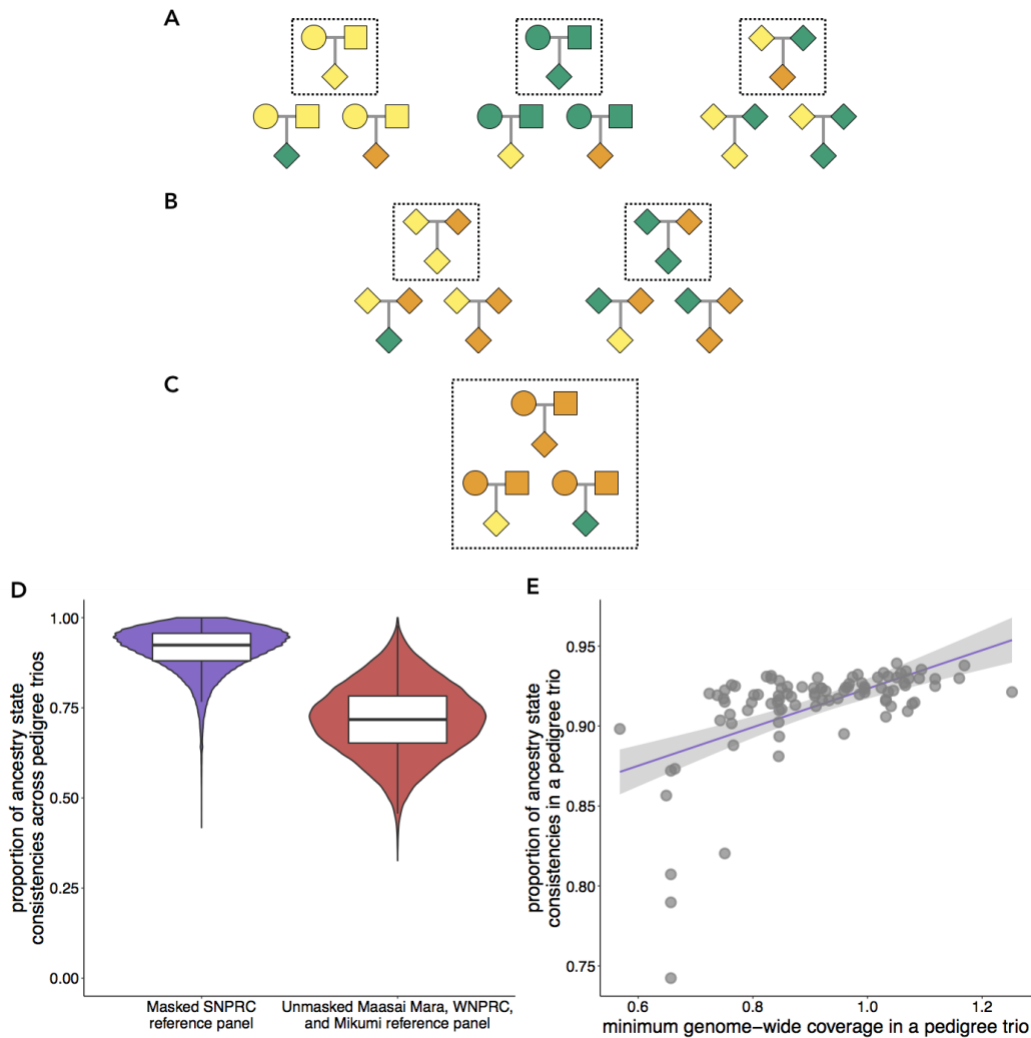
Fig. S3: Local ancestry call consistency using pedigree trios.

(A-C) Ancestry configurations for pedigree trios. Pedigree trios with internally consistent calls are shown inside black dashed boxes. Females = circles, squares = males, diamonds = males or females (i.e., the sex of the individual does not matter); yellow = homozygous yellow baboon ancestry, orange = heterozygous yellow-anubis baboon ancestry, green = homozygous anubis baboon ancestry. **(A)** Pedigree trios in which both parents are homozygotes. **(B)** Pedigree trios in which one parent has homozygous ancestry and the other parent has heterozygous ancestry. **(C)** Pedigree trios in which both parents are heterozygotes. **(D)** Proportion of ancestry state consistencies per site across pedigree trios using different reference populations for yellow baboon and anubis baboon ancestry (the SNPRC reference panels are masked for introgressed segments; see Section 6.2). Note that the SNPRC reference panel consists of only medium to high-coverage sequences, whereas the Maasai Mara/WNPRC/Mikumi reference panel contains a mix of low and high coverage sequence data. **(E)** The proportion of sites with consistent calls within trios is strongly positively correlated with the minimum genome-wide coverage among members of the trio (Pearson's r = 0.553, p = 1.07 x 10^{-8}).
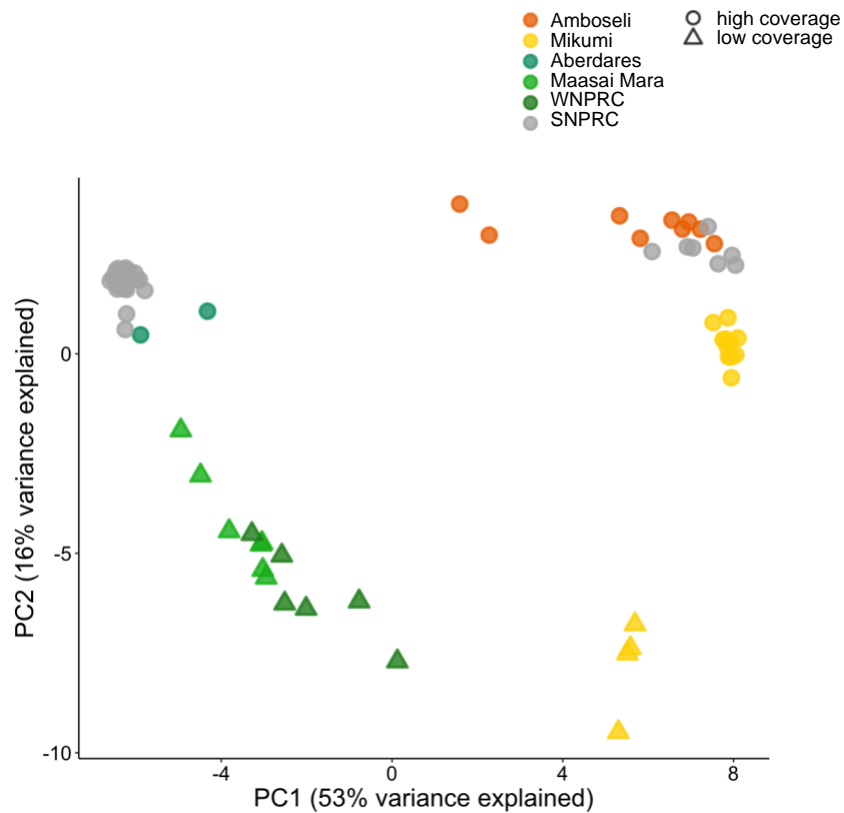
Fig. S4: The structure of genetic variation between yellow, anubis, and hybrid baboons.

Principal components analysis of genotypes from high coverage samples (SNPRC: grey circles; the Aberdares region: green circles; Mikumi: yellow circles; Amboseli: orange circles) and low coverage samples (Mikumi: yellow triangles; Maasai Mara: light green triangles; WNPRC: dark green triangles). Ancestry differences predominate on PC1, but sequencing coverage is also correlated with PC2.
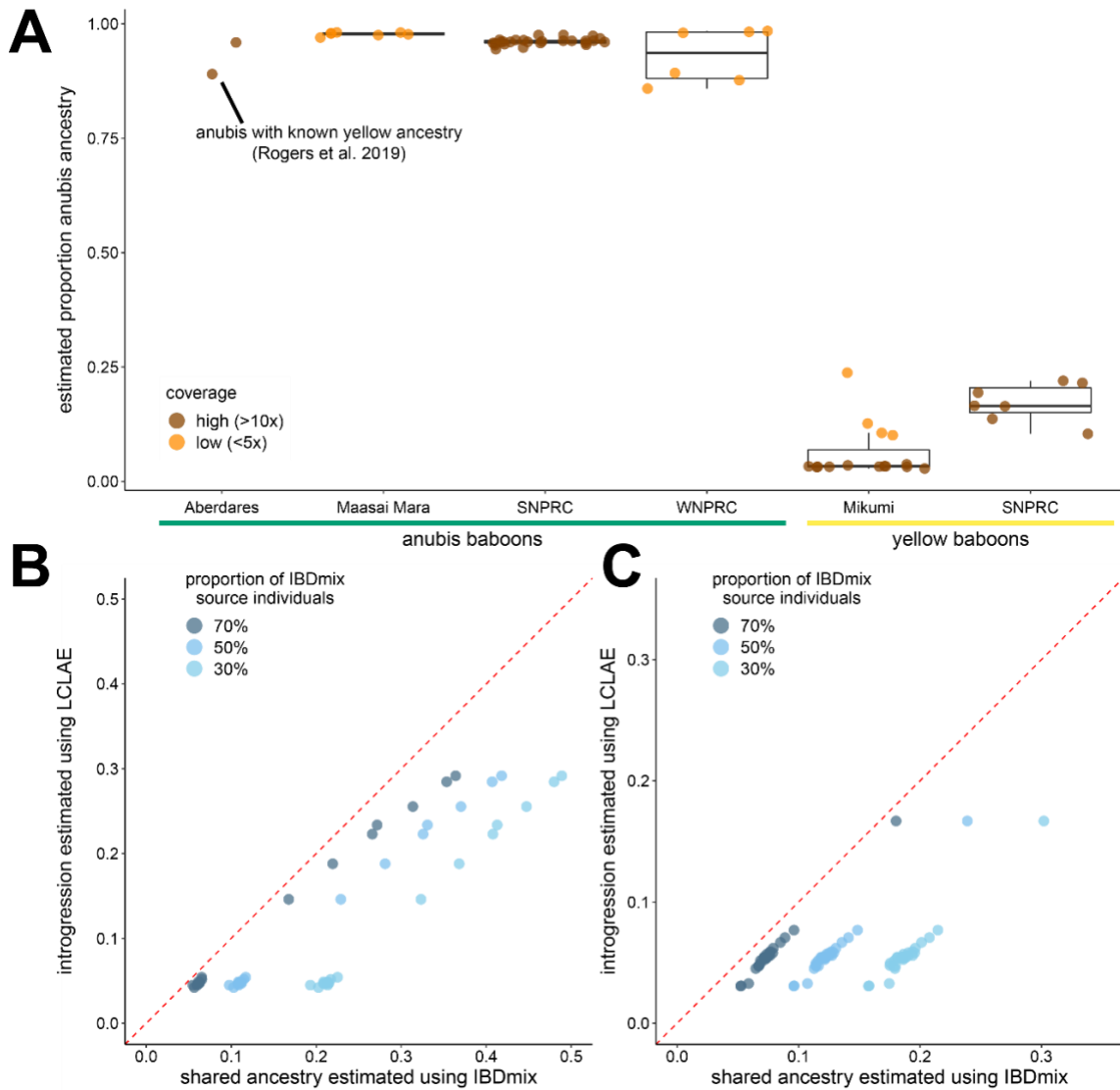
**Fig. S5: Ancestry in yellow and anubis baboons outside of Amboseli.**
(A) Genome-wide proportion of anubis ancestry inferred per baboon using LCLAE.
Estimates for high coverage individuals are shown in brown, while low coverage
individuals are shown in orange. **(B-C)** Estimates of local ancestry are strongly correlated
between IBDmix and LCLAE. (B) shows the proportion of anubis ancestry identified in
yellow baboons; (C) shows the proportion of yellow ancestry identified in anubis
baboons. Darker points represent the more conservative version of IBDmix results,
requiring that a region of the genome be identified as IBD using 70% of potential source
individuals. Successive lighter points represent increasingly liberal standards (50% or
30% of source individual agreement). Dashed red lines correspond to the y = x line,
showing that IBDmix tends to overestimate shared ancestry relative to LCLAE.
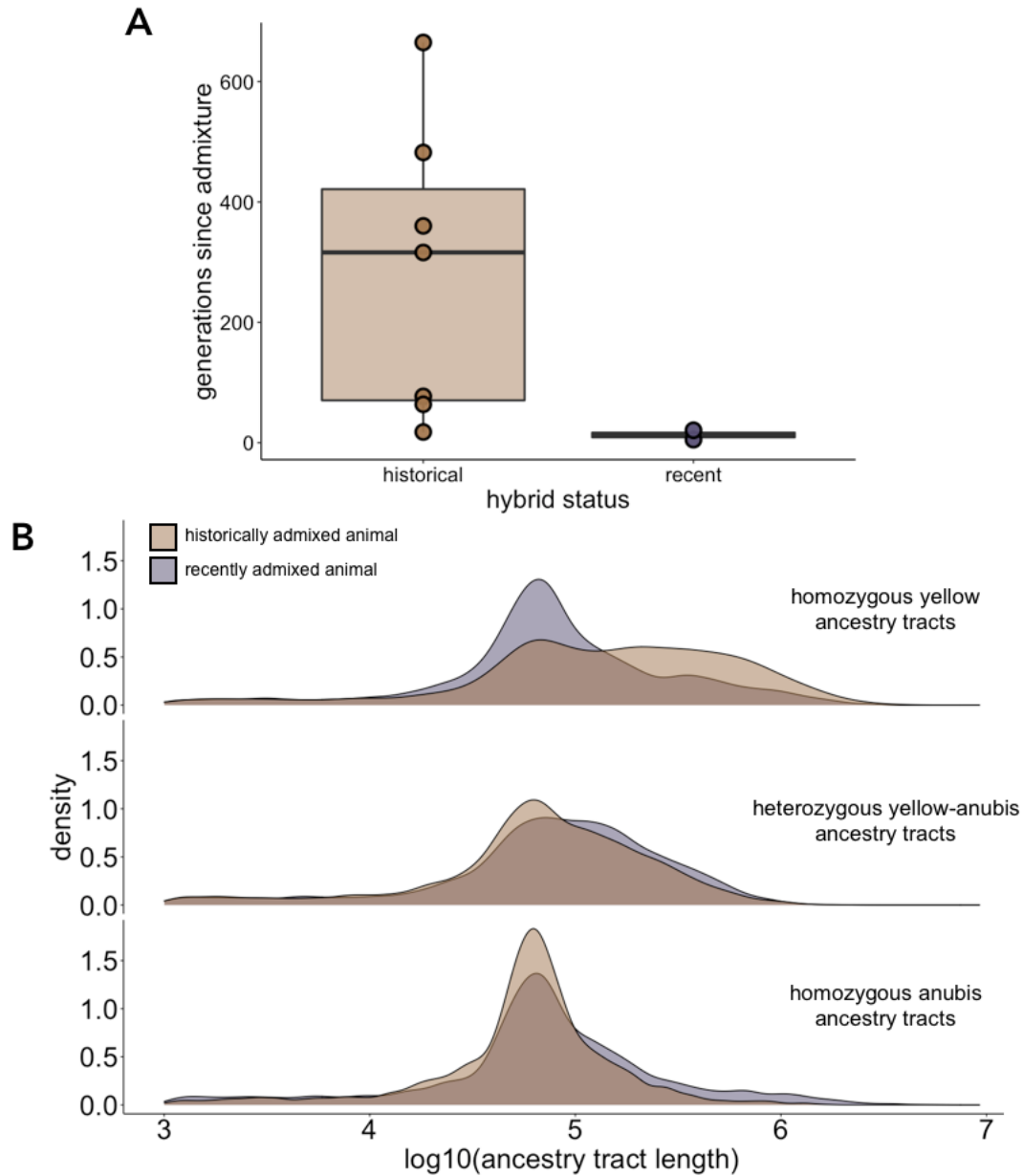
**Fig. S6: Estimated timing of admixture in Amboseli baboons.**

(A) *DATES* (*35*) estimates for the timing of admixture for Amboseli individuals sequenced at high coverage. Individuals were assigned to "recent" or "historical" hybrid categories based on whether they had a known anubis or anubis-like ancestor (Section 1.2). **(B)** The distribution of ancestry tract lengths for a recently admixed animal in (A) with an estimate of 4.99 generations since admixture (purple distributions) versus a historically admixed animal in (A) with an estimate of 316.12 generations since admixture (brown distributions).
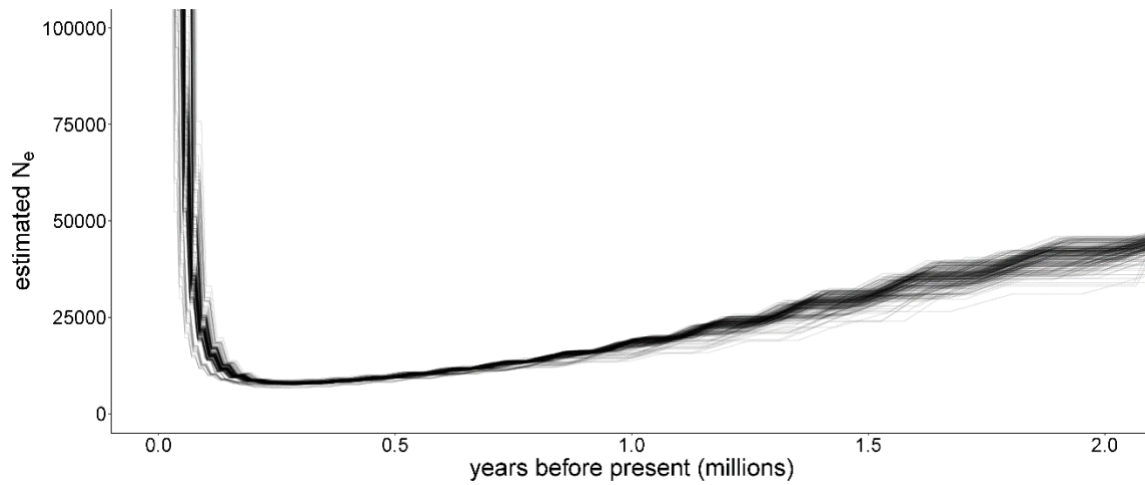
Fig. S7: Divergence of yellow and anubis baboons estimated using pseudo-hybrid PSMC.

Estimated effective population size (N_e) (y-axis) over time (x-axis), inferred from pseudo-hybrid genomes (*89*) using PSMC (*91*), for up to 2 million years in the past and 100,000 N_e. Lines correspond to results inferred for 264 yellow-anubis baboon pseudo-haploid genomes.
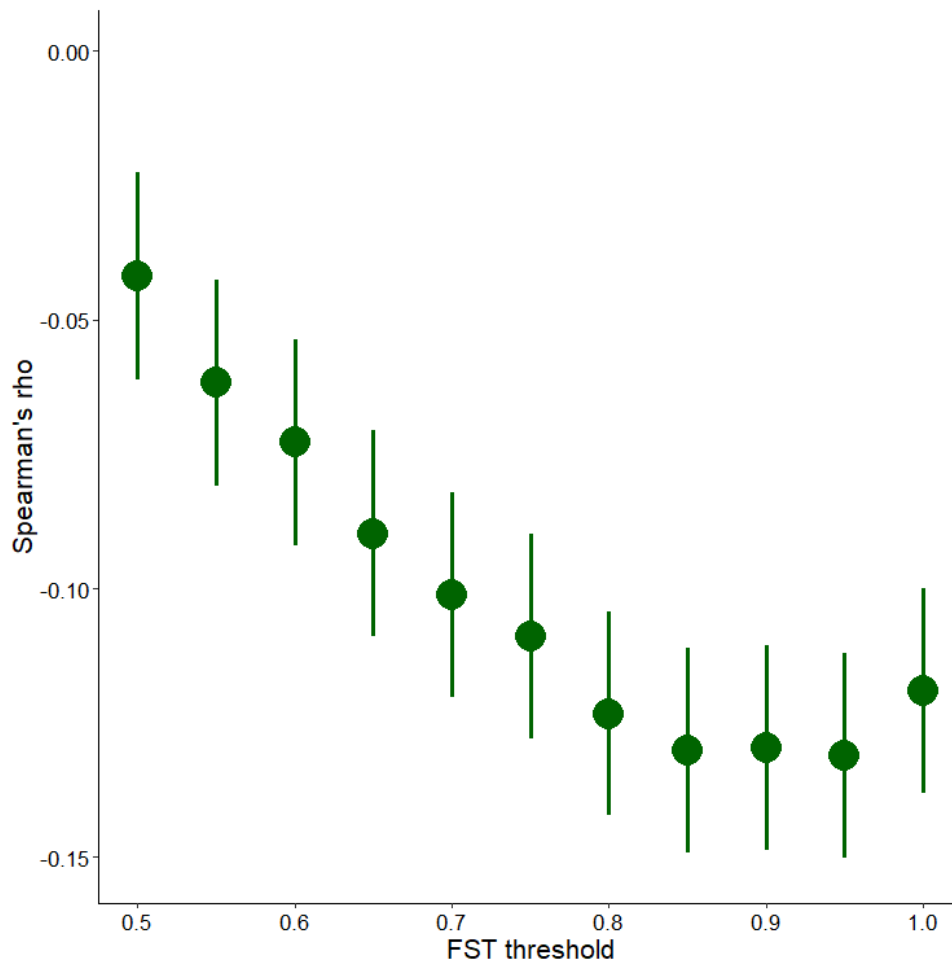
Fig. S8: Introgressed ancestry is reduced in highly diverged regions of the genome.
Spearman's rho for the correlation between introgressed ancestry and the number of highly differentiated sites in 250 kb regions of the genome, across variable minimum F_{ST} thresholds (x-axis) for defining highly differentiated sites. Error bars show the standard error of rho. For all threshold values of F_{ST}, rho is significant at $p < 2.5 \times 10^{-5}$.