Supplementary figures

The genome of the endangered Dryas monkey provides new insights into the evolutionary history of the vervets

T. van der Valk, C. M. Gonda, H. Silegowa, S. Almanza, I. Sifuentes-Romero, T. Hart, J. Hart, K. Detwiler, K. Guschanski

Corresponding authors: Tom van der Valk, tom.vandervalk@ebc.uu.se, Katerina Guschanski katerina.guschanski@ebc.uu.se

Content	Page
Table S1	2
Table S2	2
Figure S1	3
Figure S2	4
Figure S3	5
Figure S4	6
Figure S5	7
Figure S6	8
Figure S7	9
Figure S8	10
Figure S9	11
Figure S10	12
Figure S11	13
Figure S12	14

Table S1. Sample information.

Date collected	Field ID	Species	Location	GPS (decimal degrees)	Sex	Age	Country of Origin	Collector	FAU Primatology Lab ID	Yale Peabody Museum catalog number
06-Aug-14	JH023	Cercopithecus dryas	Bafundo Forest	S02.08911, E025.57543	M	Adult	DRC	John Hart	DRD14M417	YPM MAM 016890
26-Oct-14	JH024	Cercopithecus dryas	Bafundo Forest, west Kasuku River	S02.08920, E025.57500	M	Adult	DRC	John Hart	DRD14M418	-
20-Dec-17	NA	Cercopithecus dryas	Camp Ilumbe, Bafundo Forest	S02.06203, E025.58421	F	Sub-adult	DRC	Pablo Ayali	DRD17F400	YPM MAM 017066

Table S2. Primer sequences and annealing temperatures used for cytochrome B sequencing.

Primer Name	**Sequence (5' – 3')**	**Size [bp]**	**T_a[°C]**
5 overlapping Fragments			
cytbA1fw	CCACCGTTGTACTTCAACTAC	369	62
cytbA1rv	CAGGTTTTTAGGAGAAGGAATG		
cytbA2fw	GGCGCCTCCATATTTTTCATC	334	62
cytbA2rv	TTGTCTGAGTCTGATGAGATTC		
cytbB1fw	CCACCCTTTCACGATTCTTCA	382	64
cytbB1rv	ATGAGGATTGATTGATAGGAAGAGTG		
cytbB2fw	CCCTCCACACATCAAACCAG	390	62
cytbB2rv	TAGTTTACAAGACTAGTGTATTAG		
cytbB2fw4	YCCACACATCAAACCAG	322	60
cytbB2rv4	TAGAATGCCAGTTTTGGG		

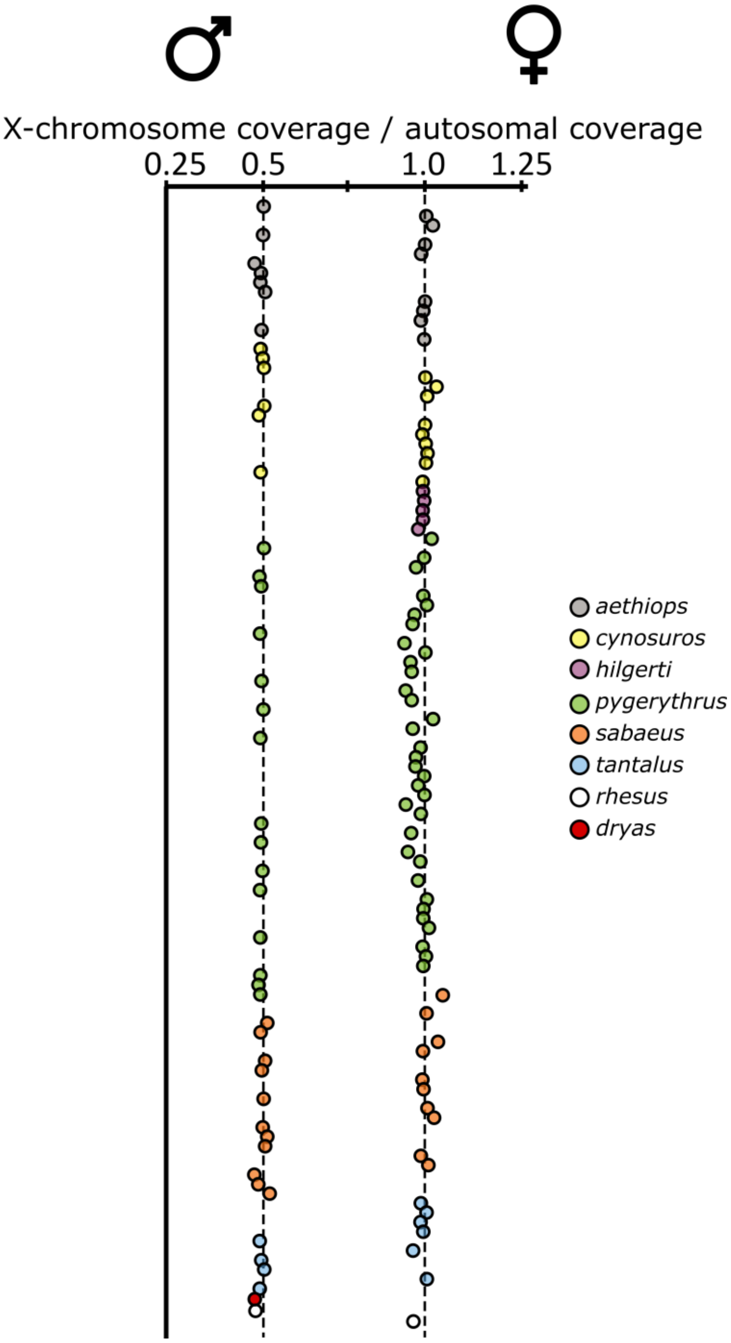

Figure S1. Sexing individuals based on the ratio of X-chromosomal to autosomal coverage.

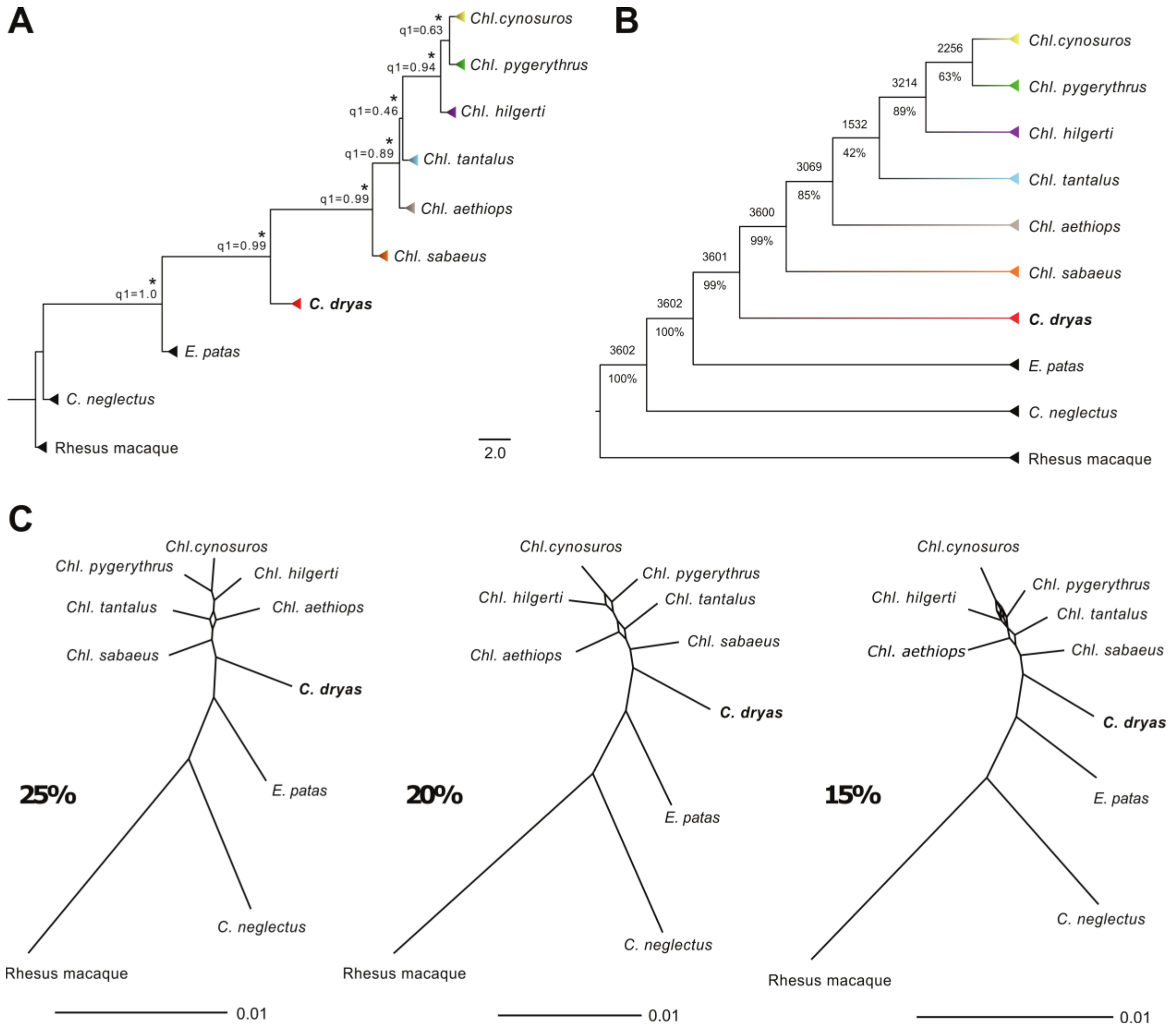

Figure S2. Phylogeny of the Dryas monkey and the vervets. (A) MSC-based species trees generated by ASTRAL using 3602 autosomal genomic windows. The tree was rooted with the rhesus macaque. Branch lengths are given in coalescent units and are an indicator of gene-tree discordance. The normalized quartet score of this topology is 0.87. Asterisk symbols at nodes indicate maximum local quartet support posterior probabilities (lpp=1.0) and *q1* values displayed at each node show the percentage of quartets in all gene trees that agree with that branch topology. (**B**) Majority-rule consensus tree obtained from 3602 autosomal gene-trees using CONSENSE in PHYLIP v3.695. The number above each branch shows the total number of trees out of all 3602 gene trees supporting the given branch and the number below corresponds to its percentage. **(C)** SplitsTree consensus networks using 3602 gene trees at median thresholds of 25%, 20% and 15%.

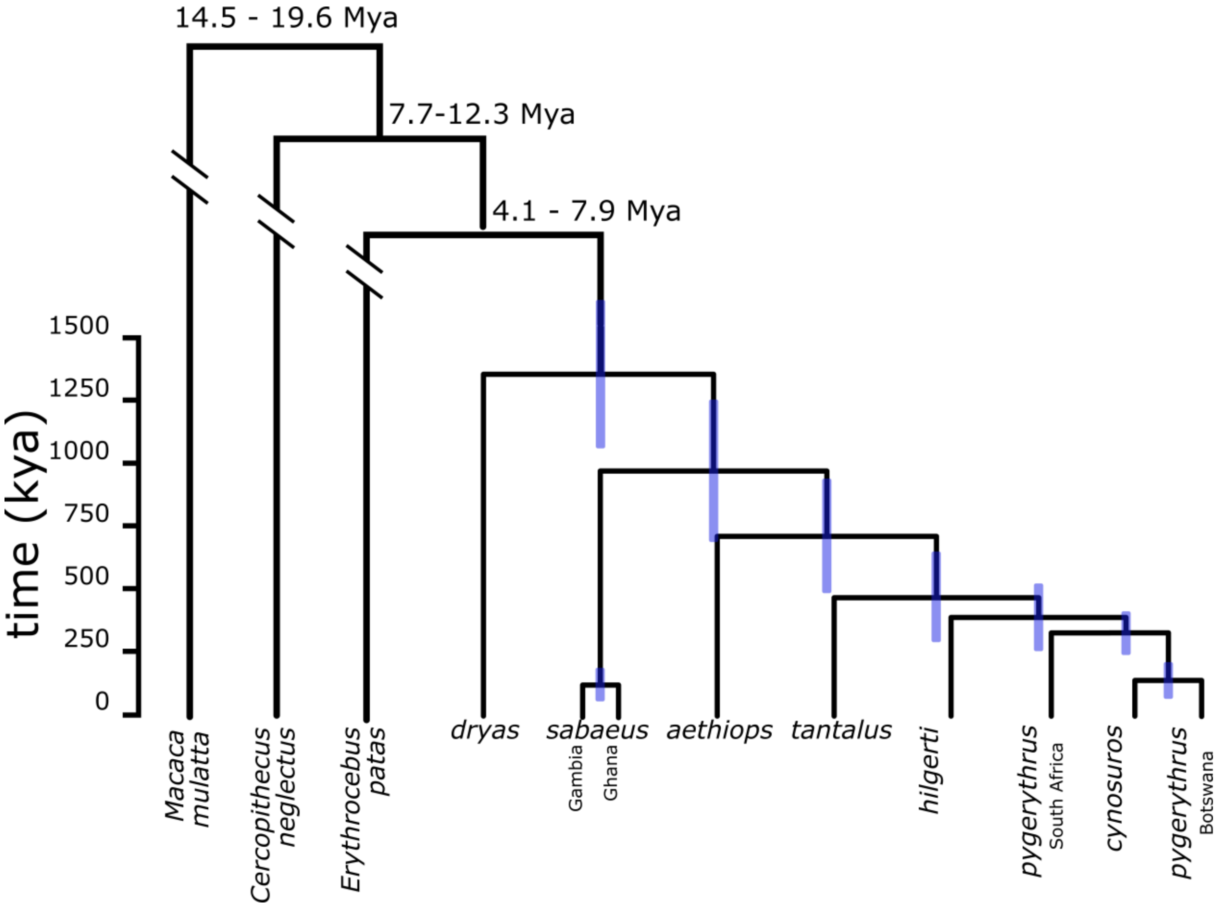

Figure S3. The autosomal consensus phylogeny and MCMC dating. Same as figure 1B but depicting 95% confidence intervals for the estimated divergence times. Confidence intervals for the oldest nodes are depicted with numbers for clarity.

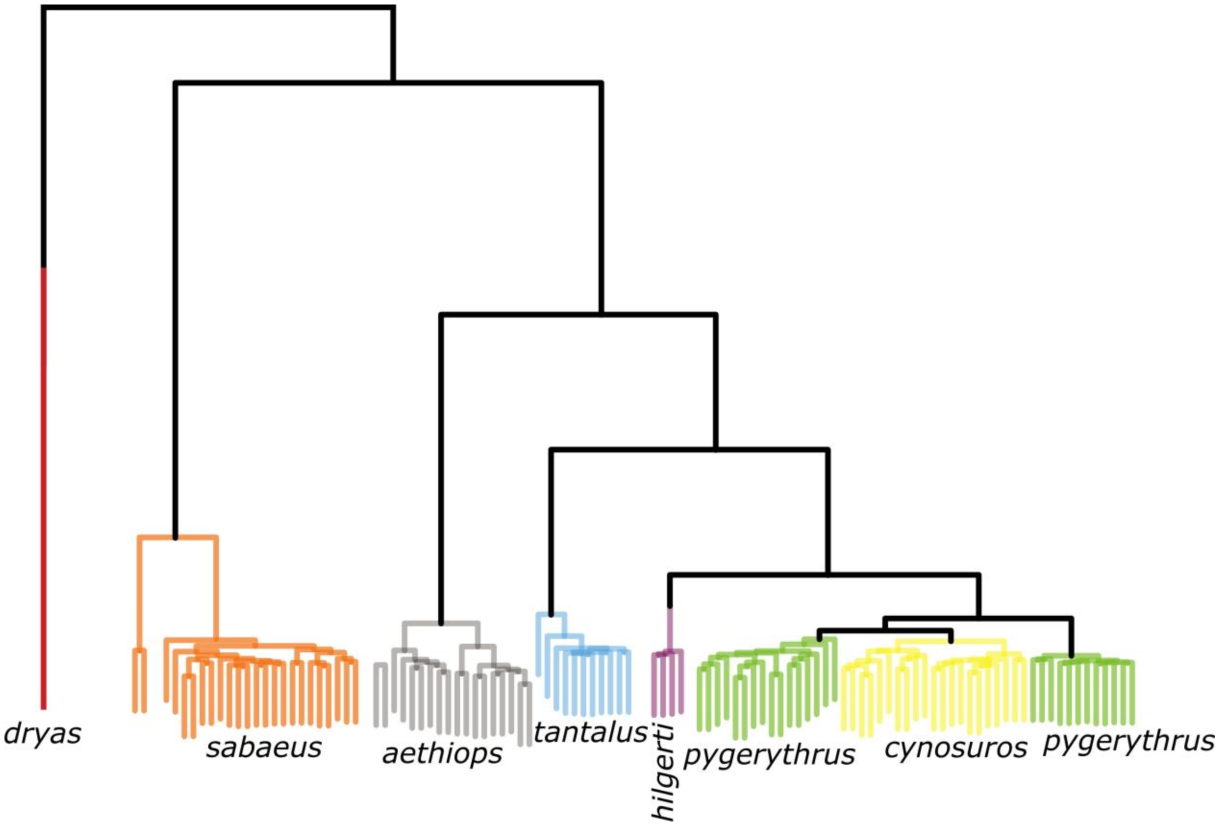

Figure S4. Evolutionary relationships of the Dryas monkey and the vervets. The tree shows the UPGMA clustering of all individuals used in this study based on all bi-allelic autosomal sites.

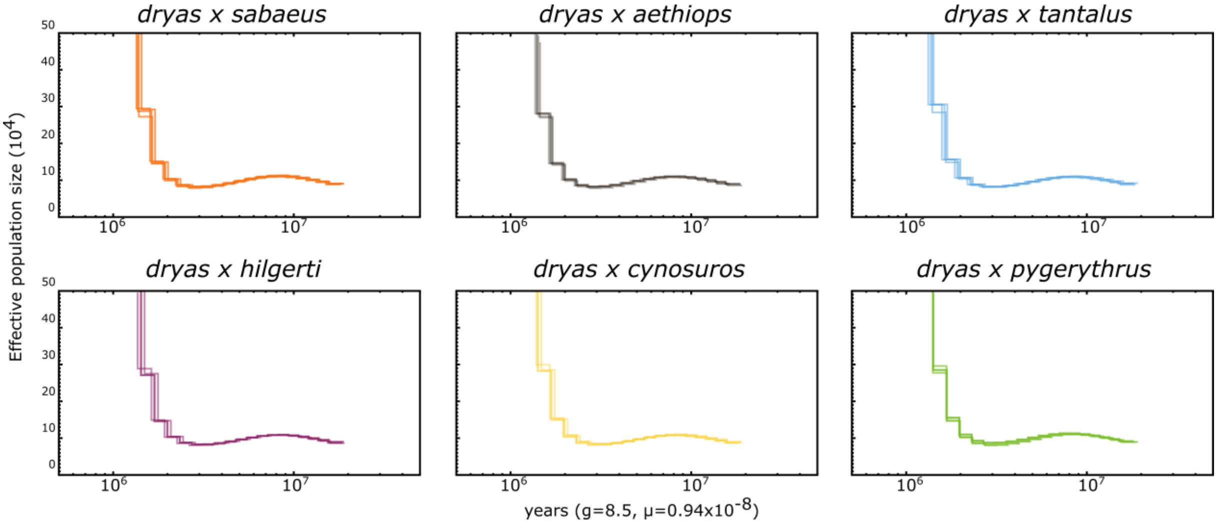

Figure S5. Artificial hybrid PSMC (hPSMC) of the Dryas monkey and the vervets. hPSMC was run on pairwise haploidized genomes for the Dryas monkey and four individuals from each vervet species assuming a generation time of 8.5 years and a mutation rate per generation of 0.94x10^{-8}. PSMC plots on artificial hybrid genomes (hPSMC) show a rapid increase in effective population size at the time when the two parental lineages diverged. A stepwise increase in Ne is an indication of a population split with ongoing gene-flow. Note that the rapid increase in effective population size occurs at the same time for each combination of Dryas monkey and vervet species, suggesting that the Dryas monkey is equidistant to each of them.

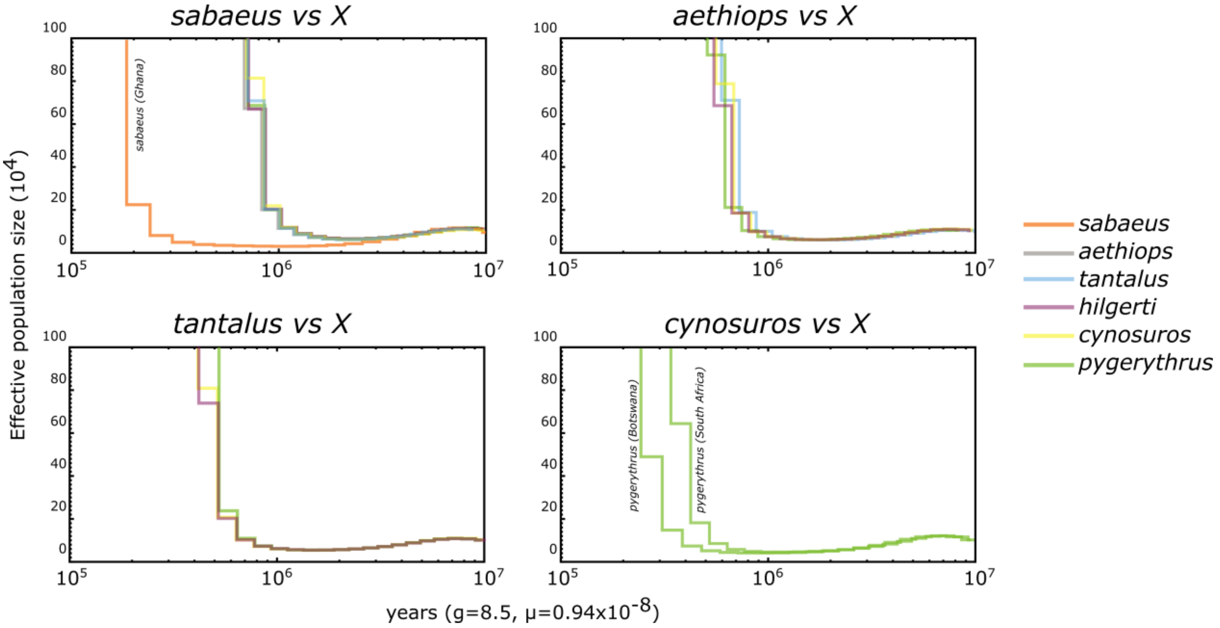

Figure S6. Artificial hybrid PSMC (hPSMC) within the vervet clade. Same as Figure S4, estimating divergence-times within the vervet genus.

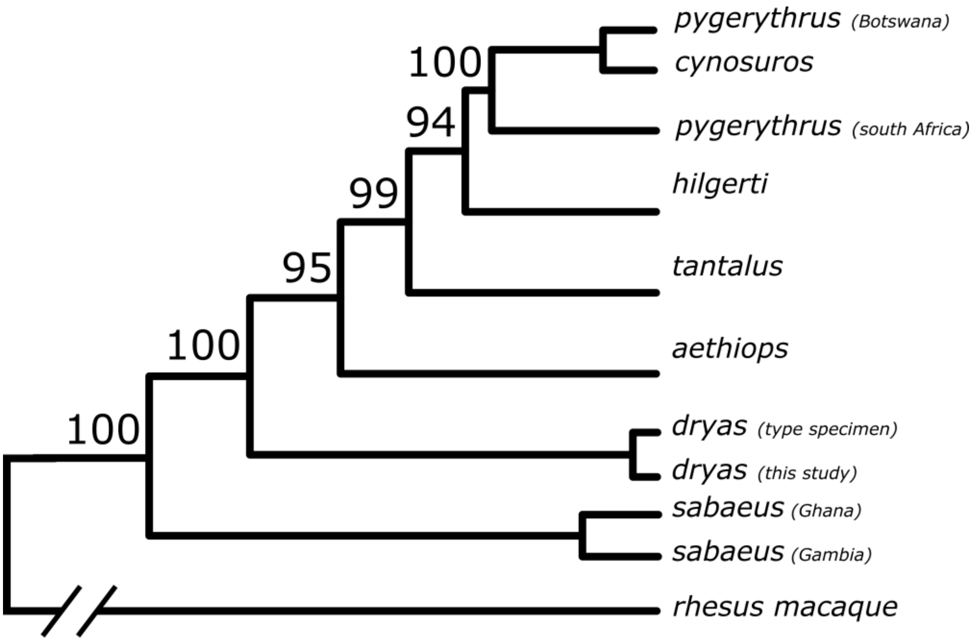

Figure S7. Maximum likelihood tree based on complete mitochondrial genomes for the vervets and the Dryas monkey (same as Figure 1C). Numbers depict bootstrap support values (1000 bootstrap replicates were run).

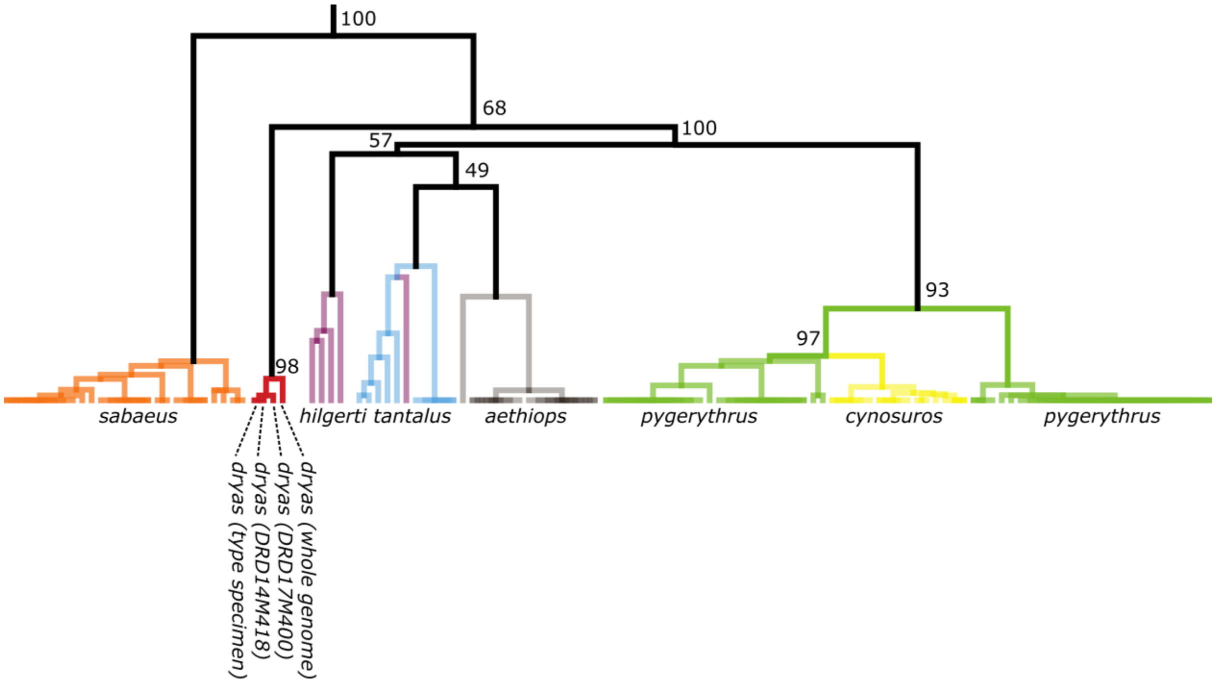

Figure S8. Maximum likelihood tree based on the cytochrome B sequence for the vervets and the Dryas monkey. Numbers depict bootstrap support values (1000 bootstrap replicates were run). Note the close clustering of all available Dryas monkey cytochrome B sequences.

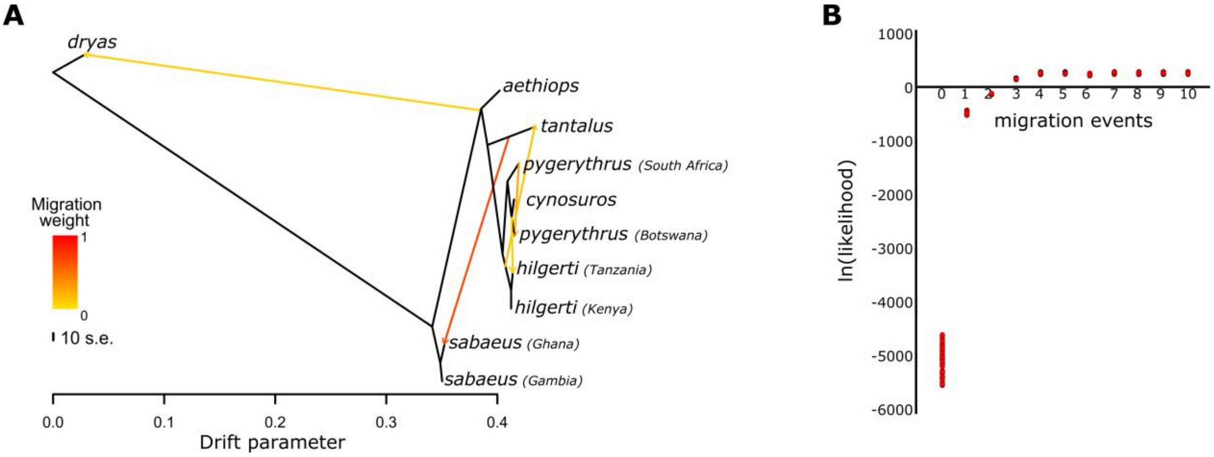

Figure S9. Model-based estimates of gene-flow. (A) TreeMix analysis for five migration events (m=5). Population allele frequencies are separated by species and geographic origin. **(B)** Likelihood support values for TreeMix models with 0 to 10 migration events respectively. After modelling five migration events, the model likelihood does not increase any further.

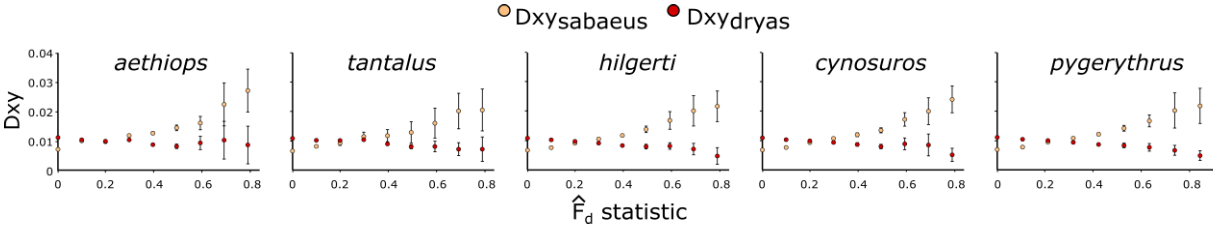

Figure S10. Identifying introgressed regions with Dxy statistic stratified by $\widehat{F}_d$. Errors bars show ±3SE. Window size = 10kb. Windows with high $\widehat{F}_d$ statistic (the excess of shared derived alleles with the Dryas monkey over that with *Chl. sabaeus*) have on average unusual large genetic distance (D_{xy}) to *Chl. sabaeus* and unusual low genetic distance to the Dryas monkey. Note that the vast majority of windows is at $\widehat{F}_d$ ~ 0 (>90%). $D_{xy\text{-sabaeus}}$ at windows with $\widehat{F}_d = 0$ (putative non-introgressed windows) is generally around 0.007, whereas $D_{xy\text{-dryas}}$ at these windows ~0.012. This is in agreement with the obtained divergence time estimates (Figure 1B) (e.g. genetic distance of the non-*sabaeus* vervets to the Dryas monkey is on average 1.75 - 2.25 times larger than to *Chl. sabaeus*). At high $\widehat{F}_d$, $D_{xy\text{-dryas}}$ is around ½-⅓ that of $\widehat{F}_d = 0$, suggesting that these windows diverged on average half to one third as long ago as the other windows (thus ~750.000 to 430.000 years ago as the most likely time period of introgression, if we assume the divergence of the Dryas monkey to the vervets to be ~1.4 Mya, as indicated by our analyses).

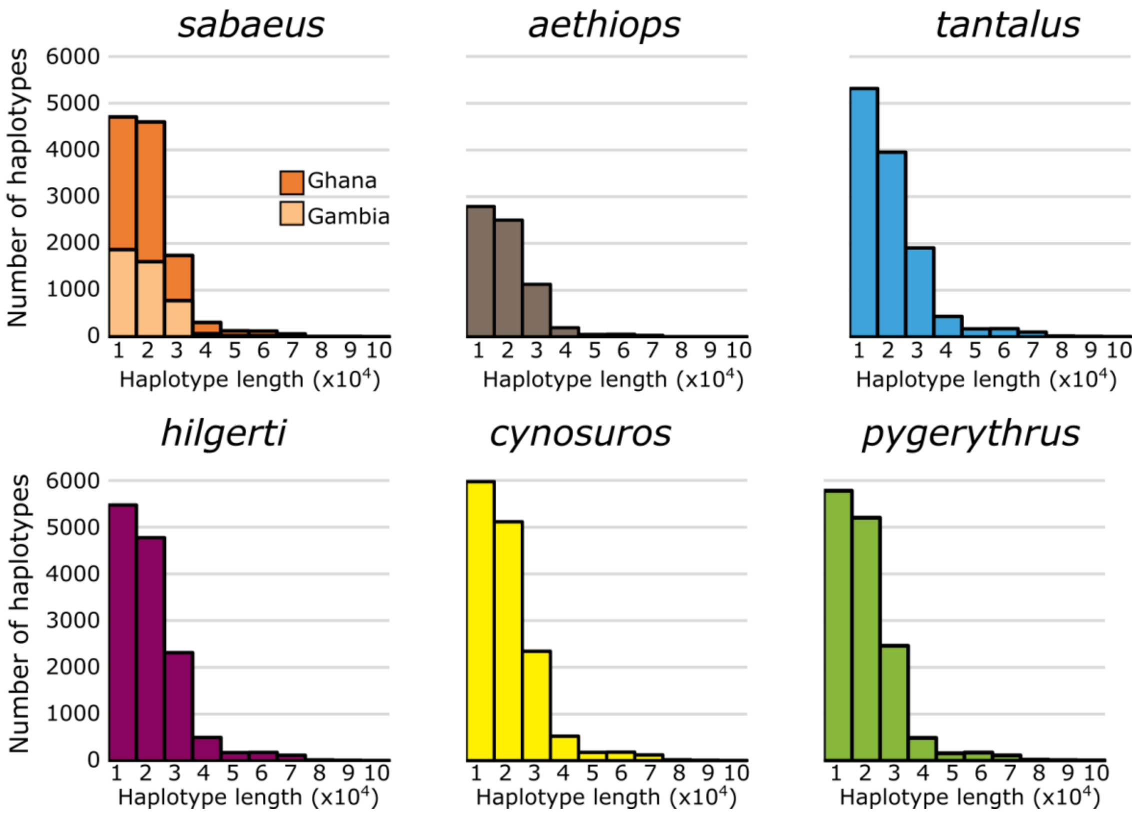

Figure S11. Length distribution of putatively introgressed haplotypes. Windows are short on average for all species. *Chl. sabaeus* individuals from Gambia have the lowest number of putatively introgressed windows, which were likely introduced into this species via secondary gene-flow with *Chl. tantalus*. Ghana individuals have a much higher number of these windows, as a result of the recent strong secondary admixture with *Chl. tantalus*.

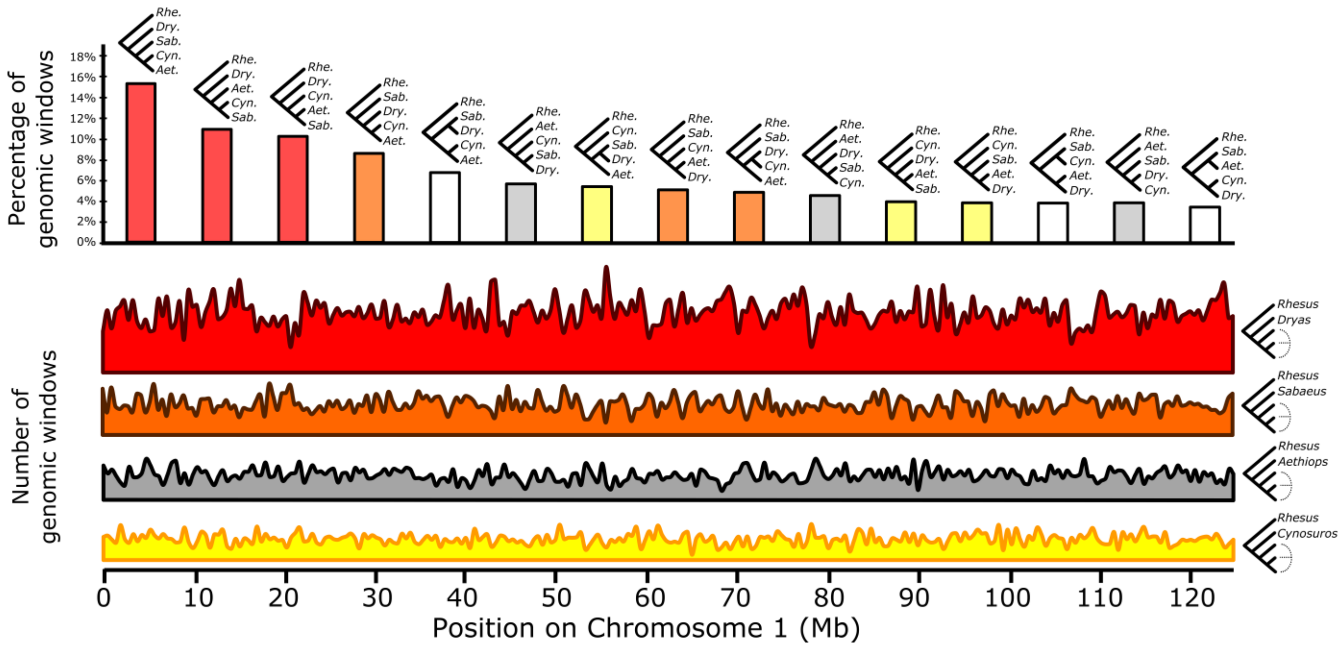

Figure S12. Exploring evolutionary relationships across the genome using topology weighting (*Twisst*). The top graph shows how often a depicted topology is supported among all non-overlapping 50 SNP windows (Rhe. = Rhesus macaque, Dry. = Dryas monkey, Sab. = *Chl. sabaeus*, Cyn. = *Chl. cynosuros*, Aet. = *Chl. aethiops*). The vast majority of windows supports the global topology (red-bars), in which the Dryas monkey is sister to all vervets. A smaller fraction of windows supports different topologies, likely to be the result of introgression and incomplete lineage sorting. Note that the second-most supported topology places the Dryas monkey inside the vervet clade with *Chl. sabaeus* as sister to all (orange bars). Plotting the occurrence of the different topologies along the genome (below), we did not identify long genomic regions supporting a contrasting topology to the majority topology (only chromosome 1 is shown here).