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

T. van der Valk[1], C. M. Gonda[1], H. Silegowa[2], S. Almanza[3], I. Sifuentes-Romero[3], T. Hart[2], J. Hart[2], K. Detwiler[3], K. Guschanski[1]

[1] Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden

[2] Frankfurt Zoological Society, TL2 Project, Kinshasa, Democratic Republic of the Congo

[3] Departments of Anthropology and Biological Sciences, Florida Atlantic University, Boca Raton, FL 33431, USA

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

Abstract

Genomic data can be a powerful tool for inferring ecology, behaviour and conservation needs of highly elusive species, particularly when other sources of information are hard to come by. Here we focus on the Dryas monkey (*Cercopithecus dryas*), an endangered primate endemic to the Congo Basin with cryptic behaviour and possibly less than 250 remaining individuals. Using whole genome sequencing data we show that the Dryas monkey represents a sister lineage to the vervets (*Chlorocebus sp.*) and has diverged from them around 1.4 million years ago with additional bi-directional gene flow ~750,000 – ~500,000 years ago that has likely involved the crossing of the Congo River. Together with evidence of gene flow across the Congo River in bonobos and okapis, our results suggest that the fluvial topology of the Congo River might have been more dynamic than previously recognised. Despite the presence of several homozygous loss-of-function mutations in genes associated with sperm mobility and immunity, we find high genetic diversity and low levels of inbreeding and genetic load in the studied Dryas monkey individual. This suggests that the current population carries sufficient genetic variability for long-term survival and might be larger than currently recognised. We thus provide an example of how genomic data can directly improve our understanding of highly elusive species.

Introduction

The Dryas monkey (*Cercopithecus dryas*) is a little-known species of guenons endemic to the Congo Basin, previously only recorded from a single location (Figure 1). The recent discovery of a second geographically distinct population in the Lomami-Lualaba interfluve area has led to the elevation of its conservation status from critically endangered to endangered, however little is known about its population size, distribution range, behaviour, ecology and evolutionary history (Hart *et al.*, 2019). Based on pelage coloration, the Dryas monkey was first classified as the Central African representative of the Diana monkey (*Cercopithecus diana*) (Schwarz, 1932). However, later examinations of the few available specimens suggested that the Dryas monkey should be classified as a unique *Cercopithecus* species (Colyn, Gautier-Hion and van den Audenaerde, 1991). More recently, Guschanski *et al.* 2013 described the mitochondrial genome sequence of the Dryas monkey type specimen, preserved at the Royal Museum for Central Africa (Tervuren, Belgium), providing the first genetic data for this species. The mitochondrial genome-based phylogeny placed the Dryas monkey within the vervet genus *Chlorocebus*, supporting previously suggested grouping based on similarities in feeding behavior, locomotion and cranial morphology (Kuroda, Kano and Muhindo, 1985; Butynski, 2013). The vervets consist of six recognised species: Chl. *sabaeus, aethiops, tantalus, hilgerti, pygerythrus* and *cynosuros* (Zinner *et al.*, 2013). They are common in savannahs and riverine forests throughout sub-Saharan Africa, as well as on several Caribbean islands, where they were introduced during the colonial times (Figure 1A). However, the Dryas monkey is geographically isolated from all vervets and has numerous highly distinct morphological characteristics (Zinner *et al.*, 2013). As vervets are characterized by a dynamic demographic history with extensive hybridisation (Svardal *et al.*, 2017), including female-mediated gene flow, transfer of mitochondrial haplotypes between species can result in discordance between the mitochondrial tree and the true species evolutionary history. Thus, the phylogenetic placement of the Dryas monkey remains uncertain.

With only two known populations and possibly fewer than 250 individuals, the Dryas monkey is of significant conservation concern (Hart *et al.*, 2019). The goal of this study was thus twofold: First, reconstruct the evolutionary and demographic history of the Dryas monkey and second, assess the long-term genetic viability of this species. To this end, we sequenced the genome of a male Dryas monkey at high coverage (33X), which represents the first genome-wide information available for this cryptic and little-known species.

59 **2.0| Methods**

60

61 **2.1| Tissue sample collection and whole genome sequencing**

62 Three Dryas monkey tissue samples were obtained from individuals from the Lomami National Park
63 buffer zone in the Democratic Republic of Congo (Figure. 1A) and exported to Florida Atlantic
64 University, United States under approved country-specific permits, where DNA was extracted using
65 the Qiaqen DNeasy Blood & Tissue kit, following the manufactures protocol. For library preparation
66 and sequencing the DNA extracts of two individuals were sent to Uppsala University, Sweden. The
67 Illumina TruSeq DNA PCR-free kit was used for standard library preparation and the sample showing
68 the highest DNA extract concentration was sequenced on one lane of the Illumina HiseqX platform (2x
69 150bp). In addition, we obtained previously published FASTQ data for all currently recognised vervet
70 species from mainland Africa (Warren *et al.*, 2015; Svardal *et al.*, 2017) and divided them into two
71 subsets: 23 *Chl. sabaeus*, 16 *Chl. aethiops*, 11 *Chl. tantalus*, 6 *Chl. hilgerti*, 16 *Chl. cynosuros* and 51 *Chl.*
72 *pygerythrus* individuals sequenced on the Hiseq2000 platform at 1.9 – 6.8 X coverage (the low-
73 coverage dataset), and one individual of each vervet species sequenced on the Genome-Analyzer II
74 platform at 7.4 – 9.8X coverage (the medium-coverage dataset). We also obtained publicly available
75 de-novo assembled genomes of two additional guenon species, the patas monkey (*Erythrocebus patas*,
76 GenBank: GCA_004027335.1) and the De Brazza's monkey (*Cercopithecus neglectus*, GenBank:
77 GCA_004027615.1), and previously published high coverage genomes (27X and 31X) of two rhesus
78 macaques (*Macaca mulatta*) to be used as outgroup (Xue *et al.*, 2016) .

79

80 **2.2| Alignment, variant detection and filtering**

81 All FASTQ data were adapter and quality trimmed using Trimmomatic on recommended settings
82 (Bolger, Lohse and Usadel, 2014) and initially aligned against the *Chlorocebus sabaeus* reference
83 genome (ChlSab1.1) (Warren *et al.*, 2015) using bwa-mem (Li, 2013) with default settings. The de-novo
84 patas and De Brazza's monkey assemblies were first converted into 100 base pair non-overlapping
85 FASTQ-reads and then aligned as above. For some analyses we noted significant reference bias arising
86 from the ingroup position of *Chlorocebus sabaeus* reference with respect to the set of species studied
87 here. As a result, FASTQ reads showing an alternative allele to the reference obtained lower mapping
88 scores than reads carrying the references allele. This bias increases with genetic distance to the
89 reference, and thus to avoid potential errors in our inferences, we additionally mapped all samples
90 against the closest available outgroup reference (*Macaca mulatta*, Mmul_10) as above. All our
91 subsequent analyses are based on the mappings to *Macaca mulatta*, unless stated otherwise.
92 Samtools was used to filter out reads below a mapping quality of 30 (phred-scale) (Li *et al.*, 2009). Next,

93 reads were realigned around indels using GATK IndelRealigner (McKenna *et al.*, 2010; DePristo *et al.*,
94 2011) and duplicates marked using Picard2.10.3 (https://broadinstitute.github.io/picard/). After these
95 filtering steps we obtained a genome-wide coverage of 33X for the Dryas monkey. Next, we called
96 single nucleotide polymorphisms (SNPs) with GATK UnifiedCaller outputting all sites (McKenna *et al.*,
97 2010; DePristo *et al.*, 2011). Raw variant calls were then hard filtered following the GATK best practices
98 (Van der Auwera *et al.*, 2013). Additionally, we removed all sites below quality 30 (phred-scale), those
99 with more than three times average genome-wide coverage across the dataset, sites for which more
100 than 75% of samples had a non-reference allele in a heterozygous state, indels and sites within highly
101 repetitive regions as identified from the repeatmask-track for the Mmul_10 reference using vcftools
102 (Danecek *et al.*, 2011).
103
104 **2.3| Cytochrome B sequencing**
105 We amplified the cytochrome B sequence (1,140 bp) of the mitochondrial genome for all three Dryas
106 monkey samples in five overlapping fragments (Haus *et al.*, 2013). PCR reactions were run in 25 µl final
107 volume containing 1U GoTaq G2 Green Master Mix, 0.33 µM of each primer (forward and reverse,
108 Table S1), 2 µl of genomic DNA and 6.5 µl ddH_20. The cycling conditions followed the procedure
109 recommended by Haus et al. 2013 with a few modifications to the annealing temperatures of each
110 primer, 94 °C for 2 minutes, followed by 40 cycles of 94 °C for 1 minute, primer-specific annealing
111 temperature (Table S1) for 1 minute, 72 °C for 1 minute, and 72 °C for 5 minutes. All PCR products
112 were checked on 2% agarose gels and then cleaned with the Wizard SV Gel and PCR Clean-up System
113 from Promega (Madison, Wisconsin) and sent off for Sanger-sequencing along with the amplification
114 primers to Molecular Cloning Laboratories (San Francisco, California). Sequence chromatograms were
115 inspected by eye for accurate base calls and assembled using Geneious R11 11.0.5.
116
117 **2.4| Autosomal phylogeny and divergence dating**
118 First we used the vervet and the Dryas monkey SNP-calls to assess the genetic similarity between all
119 individuals by applying a clustering algorithm. The R package *ape* was used to calculate pairwise
120 distances among all individuals using the Tamura and Nei 1993 model, which allows for different rates
121 of transitions and transversions, heterogeneous base frequencies, and between-site variation of the
122 substitution rate (Tamura and Nei, 1993; Paradis, Claude and Strimmer, 2004). We then used the R
123 package *phangorn* to construct a UPGMA phylogeny from the resulting pairwise distance matrix
124 (Schliep, 2011).
125 Next, we reconstructed the phylogeny of the Dryas monkey in relation to the De Brazza's monkey,
126 patas monkey and the vervets using genome-wide phylogenetic methods. First, we generated
127 haploidized sequences for one individual per species using ANGSD, by randomly selecting a single high

128 quality base call (BaseQuality ≥30, read MapQuality ≥30, max depth below 3 times the genome-wide
129 coverage) at each site in the reference genome, excluding sex-chromosomes and sites within repetitive
130 regions (Korneliussen, Albrechtsen and Nielsen, 2014). We then concatenated autosomal sequences
131 into a multi-species alignment file. Next, non-overlapping 350 kb genomic windows were extracted
132 from the alignment and filtered for missing sites using PHAST v1.4 (Hubisz, Pollard and Siepel, 2011).
133 After filtering, we excluded all windows with a length below 200 kb resulting in a final data set
134 consisting of 3602 genomic windows (mean sequence length of 298 kb ±SD 18.5 kb). Individual
135 approximately-maximum-likelihood gene trees were then generated for the set of windows using
136 FastTree2 v2.1.10 (Price, Dehal and Arkin, 2010) and the GTR model of sequence evolution. Next, we
137 constructed a coalescent species tree from the obtained gene trees with ASTRAL v5.6.2 (Zhang *et al.*,
138 2018) on default parameters. ASTRAL estimates branch length in coalescent units and uses local
139 posterior probabilities to compute branch support for the species tree topology, which gives a more
140 reliable measure of support than the standard multi-locus bootstrapping (Sayyari and Mirarab, 2016).
141 Additionally, we obtained an extended majority rule consensus tree using the CONSENSE algorithm in
142 PHYLIP v3.695 (Felsenstein, 2005). Finally, to explore phylogenetic conflict among the different gene
143 trees, we created consensus networks with SplitsTree v4 using different threshold values (15% 20%
144 and 25%) (Huson and Bryant, 2006).
145 After obtaining the consensus phylogeny, we used a Monte-Carlo-Markov-Chain (MCMC)-based
146 method to estimate divergence times between the different species. For large alignments, calculation
147 of the likelihood function during the MCMC is computationally expensive, and we therefore estimated
148 divergence times using an approximate method as implemented in the software MCMC-Tree that
149 significantly improves the speed of the MCMC calculations (Yang, 2007; Reis and Yang, 2011). For this
150 analysis we used the independent molecular clock (the rates follow a log-normal distribution) and the
151 JC69 evolutionary model with the following parameters as recommended by (Reis and Yang, 2011):
152 alpha = 0, ncatG = 5, cleandata = 1, BDparas = 1 1 0.1, kappa_gamma = 6 2, alpha_gamma = 1 1,
153 rgene_gamma = 2 20 1, sigma2_gamma = 1 10 1 , finetune = 1: .1 .1 .1 .1 .1 .1, print = 1, burnin = 2000,
154 sampfreq = 10, nsample = 20000. The divergence times were calibrated using a soft bound of 14.5–
155 19.6 Mya for the most recent common ancestor between the rhesus macaque (tribe *Papionini*) and
156 the guenons (tribe *Cercopithecini*), as recently estimated in (Reis *et al.*, 2018). We repeated the MCMC
157 run twice and checked for run convergence in R (R^2 = 0.991).
158 We also estimated the divergence times between the Dryas monkey and the vervets using Pairwise
159 Sequential Markovian Coalescent (PSMC) modelling specifically suited for low-coverage genomes
160 (Cahill *et al.*, 2016). PSMC plots on artificial hybrid genomes (hPSMC) show a rapid increase in effective
161 population size at the time when the two parental lineages diverge (Cahill *et al.*, 2016). Haploidized
162 genomic sequences of the Dryas monkey and the vervets were merged using seqtk mergefa

163 (https://github.com/lh3/seqtk.git) and we then ran hPSMC on artificial hybrids of all pairwise
164 combination between Dryas monkey and the vervets (using four different vervet genomes per species)
165 assuming a generation time of 8.5 years (Warren *et al.*, 2015) and a mutation rate per generation of
166 $0.94x10^{-8}$ (Pfeifer, 2017).
167
168 **2.5| Mitochondrial phylogeny**
169 We *de-novo* assembled the mitochondrial genomes for the Dryas monkey, the vervets, and rhesus
170 macaque with NOVOplasty (Dierckxsens, Mardulyn and Smits, 2016) with recommended settings,
171 using a K-mer size of 39 and the *Chlorocebus sabaeus* (ChlSab1.1) mitochondrial reference genome as
172 seed sequence. We also included the previously published mitochondrial genome sequence of the
173 Dryas monkey type specimen (Guschanski *et al.*, 2013). Mitochondrial genomes were aligned using
174 clustal omega with default settings (Sievers *et al.*, 2011). We then obtained a maximum-likelihood
175 phylogeny in MEGAX, using the Tamura-Nei model with uniform rates among sites and running 1000
176 bootstrap replicates (Kumar *et al.*, 2018). Next we partitioned the protein-coding genes in the
177 alignment into 1st,2nd and 3rd coding position, using the ChlSab1.1 annotation in Geneious 10.1.2
178 (https://www.geneious.com). We estimated divergence times with MCMC-tree (which allows for
179 different evolutionary rates of the partitions) using the same parameters as for the autosomal
180 divergence estimates, however with the likelihood (instead of approximate likelihood) function.
181 Finally, we also aligned the cytochrome B sequences of two additional Dryas monkey individuals to all
182 other mitochondrial genomes and reran the maximum likelihood estimates as above, this time only
183 for the cytochrome-B sequence region.
184
185 **2.6| Y-chromosome phylogeny**
186 The mammalian Y chromosome sequence is enriched for repeats and palindromes, and thus accurate
187 assembly from short-read data is challenging (Tomaszkiewicz, Medvedev and Makova, 2017; Kuderna
188 *et al.*, 2019). We therefore obtained partial Y-chromosome consensus sequences using the filtered SNP
189 calls. First, we identified all male individuals in our low-coverage dataset using the ratio of X-
190 chromosome to autosomal coverage (Figure S1). Next, GATK FastaAlternateReferenceMaker was used
191 to obtain a Y-chromosome consensus sequence for each male individual using the filtered variant calls
192 as input. We masked all sites for which at least one individual showed a heterozygous call, as these
193 represent SNP-calling errors. Additionally, we masked all repetitive regions and all sites for which one
194 or more female individuals also showed a variant call, as these regions are likely enriched for SNP-
195 errors due to mismappings. Given the scarcity of the retained genomic data (only 4% of the Y-
196 chromosome was retained), we could not use any model-based phylogenetic approaches and instead
197 constructed a UPGMA tree, as was done for the autosomal clustering, only including the filtered sites

called in all male individuals. The tree was time calibrated using relative divergence to the rhesus Y chromosome, and a rhesus macaque - guenon divergence time of 17 million years (Reis *et al.*, 2018), assuming a uniform mutation rate.

2.7| Gene flow

We performed a model-free test of unbalanced allele sharing between the vervet individuals and the Dryas monkey (D-statistic) (Green *et al.*, 2010) using all autosomal bi-allelic sites called in all individuals. First we calculated D-statistics for all individual combinations as H1,H2,H3,H4, with the Dryas monkey as H3 and the rhesus macaque as the representative of the ancestral variant H4 (we excluded sites for which the two rhesus-macaque genomes were not identical). Next, we calculated frequency-stratified D-statistics by grouping all individuals by species and calculating the frequency of shared derived alleles within each species (e.g. for each genomic site where the Dryas monkey carries the derived allele, we calculated the frequency of the derived allele in each of the vervet species), again using the Dryas monkey genome as H3 and rhesus as H4. Genome-wide D-statistic on population level were then calculated for four different bins of derived allele frequencies (<0.25, 0.25-0.5, 0.5-0.75 and >0.75).

A model-based estimate of introgression was obtained by constructing a maximum likelihood (ML) tree using TreeMix v. 1.12 (Pickrell and Pritchard, 2012), accounting for linkage disequilibrium (LD) by grouping sites in blocks of 1,000 SNPs (-k 1000). Based on the previous phylogenetic inferences, the Dryas monkey was set as root. Standard errors (-se) and bootstrap replicates (-bootstrap) were used to evaluate the confidence in the inferred tree topology and the weight of migration events. After constructing a maximum likelihood tree, migration events were added (-m) and iterated 50 times for each value of 'm' (1-10) to check for convergence in terms of the likelihood of the model as well as the explained variance following each addition of a migration event. The inferred maximum likelihood trees were visualised with the in-build TreeMix R script plotting functions.

To identify putatively introgressed regions in all vervet individuals, we performed a screen for such segments following a strategy outlined in (Martin, Davey and Jiggins, 2015). As this method is sensitive to the number of informative sites within windows, we used the mappings against the *Chlorocebus sabaeus* reference to maximise the number of informative sites (at the cost of reference bias). Briefly, in sliding windows of 10 kb we calculated $\hat{F}_d$ statistics (which is related to D-statistic but not subject to the same biases as D when calculated in sliding windows, (Martin, Davey and Jiggins, 2015)) and Dxy using all *Chl. sabaeus* individuals from Gambia as ingroup (H1) (as these showed the least amount of shared derived alleles with the Dryas monkey, see below). Genetic distance (Dxy) was calculated for each window to the Dryas monkey ($Dxy_{dryas\text{-}X}$), and to *Chl. sabaeus* ($Dxy_{sabaeus(Gambia)\text{-}X}$), where x refers to the focal individual. Next we calculated the ratio between $Dxy_{dryas\text{-}X}$ / $Dxy_{sabaeus(Gambia)\text{-}X}$ for each window. As Dxy is a measure of sequence divergence, introgressed windows between the Dryas monkey and

233 non-*sabaeus* vervets are expected to have a relatively low $Dxy_{dryas\text{-}X}$ and relatively high $Dxy_{sabaeus(Gambia)\text{-}X}$.
234 Next, windows showing an excess of shared derived alleles with the Dryas monkey ($\hat{F}_d$ value > 2
235 times standard deviation) and an unusual low genetic divergence towards the Dryas monkey (the ratio
236 of $Dxy_{Dryas\text{-}X}$ / $Dxy_{sabaeus(Gambia)\text{-}X}$ < the genome-wide average minus 2 times the standard deviation) were
237 flagged as putatively introgressed. As regions with high $\hat{F}_d$ statistics tend to cluster in regions of low
238 absolute divergence (D_{XY}) (Martin, Davey and Jiggins, 2015), some of our identified introgressed
239 regions might be false positives.
240 Additionally we used a topology weighting method, *Twisst*, for quantifying the relationships between
241 the different taxa and visualizing how these relationships change across the genome (Martin and Van
242 Belleghem, 2017). *Twisst* estimates the topology across the genome in sliding windows, where
243 consecutive blocks of contrasting topologies to the majority topology are an indication of adaptive
244 introgression. For this analysis, we first computationally phased the SNP-calls using shapeit (-phase–
245 input-vcf beagle_out.vcf–window 0.1) (Delaneau *et al.*, 2013) and subsequently obtained topologies
246 in sliding windows of 50 SNPs using phyml grouping of all samples by species, following the *Twisst*
247 guidelines (https://github.com/simonhmartin/twisst.git). As the number of topologies increases
248 exponentially with increasing number of species, running *Twisst* for many species becomes
249 computationally unfeasible. We thus ran *Twisst* on the complete dataset of inferred phyml trees from
250 the rhesus macaque, the Dryas monkey, *Chl. sabaeus*, *aethiops* and *cynosuros* samples (excluding *Chl.*
251 *pygerythrus*, *hilgerti* and *tantalus* as we did not find strong differences in introgression proportions
252 between these and *Chl. cynosuros*) using the complete method.
253
254 **2.8| Selection and gene ontology enrichment of introgressed genes**
255 Using the *Chlorocebus sabaeus* genome annotation (Warren *et al.*, 2015) we obtained for each
256 individual all genes within putatively introgressed windows as identified from $\hat{F}_d$ and Dxy statistics. A
257 gene ontology enrichment was run for all putatively introgressed genes fixed in all non-*sabaeus*
258 individuals in Blast2GO using Fisher's exact test (Gotz *et al.*, 2008). Next, for all genes we obtained
259 previously calculated selection score coefficients (XP-CLR selection scores) from Svardal *et al.* (2017).
260 These selection scores are calculated for the different vervet populations and are based on a multilocus
261 test of allele frequency differentiation, identifying regions in the genome where the change in allele
262 frequency at the locus within the vervets occurred too quickly to be explained by drift using the *XP-*
263 *CLR* method from (Chen, Patterson and Reich, 2010). Candidate genes for adaptive introgression were
264 then identified as those with high gene-selection scores (top 10% of XP-CLR scores).

265 **2.9| Demographic history**
266 To infer long-term demographic history of the studied species, we used a pairwise sequentially

267 Markovian coalescent model (PSMC) (Li and Durbin, 2011). As PSMC analysis is sensitive to the number
268 of present genomic sites (Nadachowska-Brzyska *et al.*, 2016), we used the medium coverage vervets
269 (7.4-9.8X coverage) and the Dryas monkey genome mapped to the *Chlorocebus sabeus* reference. We
270 excluded sex chromosomes, repetitive regions and all sites for which read depth was less than five and
271 higher than two times the genome-wide average. We scaled the PSMC output using a generation time
272 of 8.5 years (Warren *et al.*, 2015) and a mutation rate of 0.94×10^{-8} per site per generation (Pfeifer,
273 2017), as above. Bootstrap replicates (n=100) were performed for the high-coverage Dryas monkey
274 genome by splitting all chromosomal sequences into smaller segments using the splitfa
275 implementation in the PSMC software and then randomly sampling with replacement from these
276 fragments (Li and Durbin, 2011). Our Dryas monkey genome coverage (33X) differed strongly from that
277 of the medium-coverage vervets. As limited coverage is known to biases PSMC results (Nadachowska-
278 Brzyska *et al.*, 2016), we down-sampled the Dryas monkey genome to a similar coverage (8X) as the
279 medium-coverage vervet genomes by randomly removing 75% of reads (samtools view –s 0.25) and
280 repeated the PSMC analysis, allowing for relative comparisons between species. Demographic
281 estimates from PSMC can also be biased by admixture events between divergent populations, giving a
282 false signal of population size change (Hawks, 2017). We thus removed all putatively introgressed
283 regions (see above) from the Dryas monkey genome and re-ran the PSMC analysis.

284

285 **2.10| Heterozygosity and inbreeding**

286 We measured genome-wide autosomal heterozygosity for all individuals with average genome
287 coverage > 3X using realSFS as implemented in ANGSD, considering only uniquely mapping reads (-
288 uniqueOnly 1) and bases with quality score above 19 (-minQ 20) (Fumagalli *et al.*, 2013; Korneliussen,
289 Albrechtsen and Nielsen, 2014). ANGSD uses genotype-likelihoods, rather than variant calls, allowing
290 for the incorporation of statistical uncertainty in low-coverage data and shows high accuracy in
291 estimating heterozygosity for genomes above 3X coverage (Fumagalli, 2013; van der Valk *et al.*, 2019).
292 Next, we used PLINK1.9 (Purcell *et al.*, 2007) to identify stretches of the genome in complete
293 homozygosity (runs of homozygosity: ROH) for all individuals with average genome coverage > 3X. To
294 this end, we ran sliding windows of 50 SNPs on the VCF files of all included genomes, requiring at least
295 one SNP per 50kb. In each individual genome, we allowed for a maximum of one heterozygous and
296 five missing calls per window before we considered the ROH to be broken.

297

298 **2.11| Genetic load**

299 To estimate genetic load in the Dryas monkey and vervets we used the mappings and genome
300 annotation of the *Chlorocebus sabaeus*. The variant effect predictor tool (McLaren *et al.*, 2016) was

301 used to identify loss-of-function mutations (transcript ablation, splice donor variant, splice acceptor
302 variant, stop gained, frameshift variant, inframe insertion, inframe deletion, and splice region variant),
303 missense and synonymous mutations on the filtered SNP calls. As an indication of mutational load, for
304 each individual we counted the number of genes containing one or more loss-of-function and the total
305 number of missense mutations divided by the number of synonymous mutations (dN/dS) (Fay, Wyckoff
306 and Wu, 2001). We excluded all missense mutations within genes containing a loss-of-function
307 mutation, as these are expected to behave neutrally. Dividing by the number of synonymous mutations
308 mitigates species-specific biases, such as mapping bias due to the fact that the reference genome was
309 derived from a *Chl. sabaeus* individual, coverage differences, and mutation rate (Xue *et al.*, 2015).

310 **3.0 | Results and discussion**

311

312 **3.1| The dynamic demographic history of the Dryas monkey and the vervets**

313 In this study we present the first genome sequence of the Dryas monkey, and show that it is a sister
314 lineage to all vervets (Figure 1B, S2, S3). Multiple phylogenomic approaches (MSC, Maximum likelihood
315 and Bayesian) unambiguously support the same tree topology, thus contradicting the suggested
316 placement within the *Chlorocebus* genus as inferred from the mitochondrial data (Figure 1C,
317 Guschanski *et al.*, 2013). After analysing 3602 gene trees from autosomal genomic windows, we
318 obtained a multi-species coalescent tree with maximum support values (lpp=1.0) for all nodes (Figure
319 S2A). Our topology is consistent with the vervet phylogeny previously reported by Warren *et al.* 2015.
320 Although a majority-rule consensus tree (Figure S2B) and network analyses at different threshold
321 values showed some poorly resolved nodes within the vervet clade (Figure S2C), the position of Dryas
322 monkey as sister lineage to all vervets remains unambiguous (Figure S2). By using an approximate
323 likelihood MCMC method on all autosomes we estimate that the Dryas monkey diverged from the
324 vervets around 1.4 million years ago, whereas the estimates of divergence times to both the genus
325 *Cercopithecus* as exemplified by the De Brazza's monkey (~10.2Mya) and the genus *Erythrocebus*
326 represented by the patas monkey (~5.8 Mya) are much older (Figure 1B, S3). Artificial hybrid PSMC
327 (hPSMC) analysis, which is based on mutation rate rather than calibration nodes, suggests a divergence
328 with ongoing gene-flow between the Dryas monkey and the vervets starting at ~2Mya ago with the
329 final separation of the two lineages around 1.5 Mya (Figure S5). Both the MCMC and hPSMC analyses
330 suggest that the radiation within the vervets started ca. 960 thousand years ago (Figure 1B, S6).
331 Although our inferred divergence times are sensitive to both node calibration used in the MCMC and
332 the mutation rate used for the hPSMC, they unambiguously support the sister relationship between
333 the Dryas monkey and the *Chlorocebus* genus. Based on our analyses with only two representatives of
334 the genus *Cercopithecus* (the Dryas monkey and de Brazza's monkey), this genus appears paraphyletic.
335 More guenon genomes will be needed to disentangle their complex evolutionary history, which may
336 call for reconsideration of the taxonomic classification of the Dryas monkey.

337 The divergence times within the vervets inferred by us are more ancient than reported by Warren *et*
338 *al.* (2015). This is likely explained by the use of a more ancient date for the divergence between the
339 rhesus macaque and vervets as calibration node (14.5-19.6 Mya in this study versus 12 Mya in Warren
340 *et al.* 2015), the mapping against an outgroup (*Macaca mulatta*) in contrast to Warren et. al 2015 who
341 used an ingroup reference (*Chlorocebus sabaeus*), which likely artificially increased the similarity to
342 *Chl. sabaeus* thus lowering the estimated divergence time, and the used methods (MCMC and hPSMC
343 in this study versus pairwise differences in Warren et. al 2015). The Y-chromosome-based phylogeny
344 shows the same topology and similar divergence time estimates as the autosomal tree (Figure 1D).

345 This stands in stark contrast to the inferences based on mitochondrial sequences, which show that all
346 four Dryas monkey individuals with available mtDNA sequence data are nested within the vervet genus
347 *Chlorocebus* (Figure 1C, S7, S8).
348 The discrepancies in tree topologies derived from genomic regions with different inheritance modes
349 (autosomal and mitochondrial) and the known history of introgression among the vervets (Svardal *et*
350 *al.*, 2017) suggest a possible role of gene flow in shaping the evolutionary history of the Dryas monkey.
351 Therefore, we explored if ancient admixture events can resolve the observed phylogenetic discordance
352 between the autosomal and mitochondrial data. We found that *Chl. sabaeus* individuals from Gambia
353 share significantly fewer derived alleles with the Dryas monkey (D-statistic) than all other vervets
354 (Figure 2A). As derived alleles should be approximately equally frequent in all species under the
355 scenario of incomplete lineage sorting without additional gene flow (Green *et al.*, 2010), such a pattern
356 strongly suggests that alleles were exchanged between the Dryas monkey and the vervets excluding
357 *Chl. sabaeus* that has split off from the others ~960 kya (Figure 1B). *Chl. aethiops* individuals share
358 significantly fewer derived alleles with the Dryas monkey than *Chl. tantalus, hilgerti, cynosuros*, and
359 *pygerythrus* (Figure 2A), suggesting that gene flow likely occurred for an extended period of time, at
360 least until after the separation of *Chl. aethiops* from the common ancestor of the other vervets
361 (~710kya, Figure 1B). As *Chl. tantalus, hilgerti, cynosuros* and *pygerythrus* vervets share a similar
362 amount of derived alleles with the Dryas monkey, gene flow most likely ended before this group
363 radiated (~460kya, Figure 1B). The small observed differences in the D-statistic between *Chl. tantalus,*
364 *hilgerti, cynosuros* and *pygerythrus* (±12%, Figure 2A) could be the result of drift and selection due to
365 population size differences among these vervet species or hint at ancient substructure within these
366 species (Slatkin and Pollack, 2008; Svardal *et al.*, 2017). The inferred history of gene flow is concordant
367 with the mitochondrial phylogeny, which suggests that the Dryas monkey mitochondrial genome was
368 introgressed from the common ancestor of all non-*sabaeus* vervets ~810 kya (Figure 1B and 1C). Gene
369 flow between the common ancestor of the non-*sabaeus* vervets and the Dryas monkey is also
370 supported by TreeMix analyses (Figure S9), but we note that these inferences have to be interpreted
371 with caution, as this model-based approach relies on accurate allele frequency estimates, which are
372 absent for the Dryas monkey population, as it is represented by a single individual.
373 In contrast to the *Chl. sabaeus* individuals from Gambia, *Chl. sabaeus* vervets from Ghana also carry an
374 excess of shared derived alleles with the Dryas monkey (Figure 2A). The Ghanese *Chl. sabaeus*
375 population recently hybridised with *Chl. tantalus* and thus a large proportion of their genome (~15%)
376 is of recent *Chl. tantalus* ancestry (Figure 1B) (Svardal *et al.*, 2017). As *Chl. tantalus* individuals carry
377 many shared derived alleles with the Dryas monkey, this secondary introgression event likely led to
378 the introduction of Dryas monkey alleles into the Ghanese *Chl. sabaeus,* explaining the high D-statistics
379 in this population.

380 Next, we obtained approximations of the directionality of gene flow using frequency-stratified D-
381 statistics as in (de Manuel *et al.*, 2016). We found that the vervet populations carry derived alleles
382 shared with the Dryas monkey at either low or high frequency, but few such alleles are found at
383 intermediate frequencies (Figure 2B). High frequency alleles in the donor population are more likely
384 to be introgressed and are subsequently present at low frequency in the recipient population
385 (Kuhlwilm *et al.*, 2016). Therefore, our observation is consistent with bi-directional gene flow between
386 the Dryas monkey and the non-*sabaeus* vervets. The general direction of the gene flow appears to
387 have been dominated by the introgression from the Dryas monkey into the non-*sabaeus* vervets, as
388 we observe an overall higher proportion of low-frequency shared derived alleles in the vervets (Figure
389 2B). This difference is particularly pronounced in *Chl. aethiops*, suggesting that the gene flow was
390 primarily from the Dryas monkey into the vervets before *Chl. aethiops* separated from the common
391 ancestor of *Chl. tantalus, hilgerti, cynosuros* and *pygerythrus*. After this split, gene flow likely became
392 more bi-directional with increased proportion of introgression events into the Dryas monkey, as
393 evidenced by the presence of high frequency putatively introgressed alleles in *Chl. tantalus, hilgerti,*
394 *cynosuros* and *pygerythrus*. An alternative explanation is that the observed allele frequencies are
395 driven by selection, as introgressed alleles might be on average selected against (Schumer *et al.*, 2018).
396 It is also noteworthy that all four Dryas monkey individuals for which mitochondrial sequence data are
397 available carry the vervet-like mitochondrial haplotype (Figure S8), which must have been introduced
398 into the population through female-mediated gene flow. Given that the three Dryas monkeys in this
399 study come from a different population than the Dryas monkey museum type specimen, this vervet-
400 like haplotype is most likely fixed (or present at high frequency) in the *C. dryas* population.
401 The putatively introgressed alleles in the *Chl. sabaeus* population from Ghana are found at
402 intermediate frequency (> 0.25 and < 0.50) (Figure 2B), which is in agreement with the indirect
403 introduction of these alleles through recent introgression from *Chl. tantalus*. As the *Chl. tantalus*
404 population carries derived alleles at high and low frequency (Figure 2B), the Ghanese *Chl. sabaeus*
405 population received a mixture of both high and low frequency alleles, resulting in the observed
406 intermediate frequency.
407 Using approaches that are relatively insensitive to demographic processes (e.g. genetic drift and
408 changes in effective population size), we obtained strong support for the presence of gene flow
409 between the Dryas monkey and the vervets, but we caution that they may incorrectly infer gene flow
410 in situations with ancestral subdivision (Slatkin and Pollack, 2008). However, such ancestral population
411 structure would have to persist over an extended period of time, encompassing multiple speciation
412 events. Furthermore, our inferences of gene flow are supported by the discordance between the
413 nuclear and mitochondrial phylogenies. Thus, gene flow seems to be the most parsimonious
414 explanation.

415 **3.2| Identifying introgressed regions and inferring their functional significance**

416 Using D_{xy} and $\hat{F}_d$ statistic (Martin, Davey and Jiggins, 2015) we identified putatively introgressed regions
417 in sliding windows of 10.000 bp for each individual. As expected under the scenario of secondary gene
418 flow, windows containing an excess of shared derived alleles with the Dryas monkey have low genetic
419 divergence (D_{xy}) to the Dryas monkey and high genetic divergence to *Chl. sabaeus* (Figure S10) (de
420 Manuel *et al.*, 2016). Summing over all putatively introgressed windows, we roughly estimate that
421 0.4% - 0.9% in Gambia *Chl. sabaeus*, 1.6 – 2.4% in *Chl. aethiops* and 2.7% - 4.8% of the genome in the
422 other vervets show as signature of introgression with the Dryas monkey. However we note that we
423 can only identify the excess of Dryas monkey ancestry over that present in the *Chl. sabaeus* population.
424 These estimates are thus likely a lower bound, as the *Chl. sabaeus* population might carry some Dryas
425 monkey ancestry due to secondary gene-flow with other vervet species (Figure 1B, Svardal *et al.*, 2017).
426 We estimate that putatively introgressed haplotypes average less than 10,000 bp and identify
427 putatively introgressed blocks of up to 180 kb (Figure S11). However, we note that we cannot estimate
428 the length of introgressed blocks below the length of 10,000 bp, as such short windows do not provide
429 sufficient number of informative sites. The similar length of putatively introgressed haplotypes in all
430 non-*sabaeus* vervet species strongly supports that gene flow occurred in their common ancestor. The
431 putatively introgressed haplotypes into the Ghanese and (to a lesser extent) Gambian *Chl. sabaeus*
432 populations were later introduced during secondary gene flow with *Chl. tantalus*, as an independent
433 recent gene flow event from the Dryas monkey into these *Chl.* sabaeus individuals would have resulted
434 in significantly longer haplotypes (Liang and Nielsen, 2014). However, we caution that our ability to
435 accurately identify the haplotype lengths is low given the short length of the introgressed haplotypes
436 and a single available Dryas monkey genome.

437 Interestingly, the proposed gene-flow between bonobos (*Pan paniscus*) and non-western chimpanzees
438 (*Pan troglodytes*), which have overlapping distribution range with the Dryas monkey and the vervets,
439 respectively, possibly occurred around a similar time period, (~500.000 years ago, de Manuel *et al.*,
440 2016). It is noteworthy that the gene-flow that occurred between the Dryas monkey and the vervets
441 most likely involved the crossing of the Congo River (Figure 1), previously thought to be an
442 impenetrable barrier for mammals (Colyn, 1987; Colyn, Gautier-Hion and Verheyen, 1991; Colyn and
443 Deleporte, 2004; Eriksson *et al.*, 2004). A similar scenario of cross-Congo River gene flow was proposed
444 for the okapi populations, and bonobos-chimpanzees (Stanton *et al.*, 2014; de Manuel *et al.*, 2016),
445 suggesting that the fluvial topology of the Congo River and the geology within the Congo basin might
446 have been more dynamic 750,000 - 500,000 years ago than previously recognised (Beadle, 1981;
447 Stankiewicz and de Wit, 2006).

448 Having identified introgressed regions, we explored if they may carry functional significance. First, we
449 used *Twisst* to estimate the most likely topology along short sliding windows (50 SNPs), where multiple

450 consecutive windows showing a contrasting topology to the majority topology can indicate adaptive
451 introgression (Martin and Van Belleghem, 2017). Although this analysis supported our inferred species
452 topology with additionally high support of introgression between the Dryas monkey and the vervets,
453 we did not detect strong signals of long introgressed blocks, likely due to the very ancient timing of
454 gene flow (Figure S11). As *Twisst* does not distinguish between incomplete lineage sorting and
455 introgression for very short windows, we focused on introgressed regions identified with $\hat{F}_d$/Dxy
456 statistics. We find that genes previously identified to be under strong selection within vervets (top 10%
457 of genes under the strongest selective pressure (Svardal *et al.*, 2017)) are less often introgressed with
458 the Dryas monkey (average introgression frequency 0.021) than genes that did not experience strong
459 selection (bottom 10%; average allele frequency 0.029). This may indicate weak selection against
460 introgressed genes on average, which may be deleterious in the non-host background, a pattern also
461 observed for Neanderthal genes in *Homo sapiens* and bonobo genes in the chimpanzee genetic
462 background (Sankararaman *et al.*, 2014; Schumer *et al.*, 2018). However, we can not exclude that the
463 difference in these gene frequencies is due to stochastic events or drift. To identify genes with adaptive
464 functions, we focused on 109 putatively introgressed genes that are fixed in all non-*sabaeus* vervets.
465 Gene ontology analysis did not reveal significant enrichment for any functional category (FDR=1).
466 Vervets are the natural host of the simian immunodeficiency virus (SIV) and the genes under strongest
467 selection in the vervets are related to immunity against this virus (Svardal *et al.*, 2017). We find
468 POU2F1, AEBP2, and PDCD6IP among fixed putatively introgressed genes in all non-*sabaeus*
469 individuals. POU2F1 is a member of the pathway involved in the formation of the HIV-1 elongation
470 complex in humans (Sturm *et al.*, 1993), AEBP2 is a RNA polymerase II repressor (Kim, Kang and Kim,
471 2009) known to interact with viral transcription (Zhou and Rana, 2002; Debaisieux *et al.*, 2012), and
472 PDCD6IP is involved in virus budding of the human immunodeficiency and other lentiviruses (Strack *et*
473 *al.*, 2003; von Schwedler *et al.*, 2003). However, the SIV resistance related genes that experienced the
474 strongest selection in vervets (e.g. RANBP3, NFIX, CD68, FXR2 and KDM6B) do not show a signal of
475 introgression between vervets and the Dryas monkey. Therefore, while adaptive importance can be
476 plausible for some of the introgressed loci, it does not appear to be a strong driver for retaining
477 particular gene classes.
478
479 **3.3| Genomic view on conservation of the endangered Dryas monkey**
480 The Dryas monkey is considered the only representative of the Dryas species group (Grubb *et al.*, 2003)
481 and is listed as endangered in the IUCN Red List due to its small population size of ca. 250 individuals
482 (Hart *et al.*, 2019). We therefore used demographic modelling and genome-wide measures of
483 heterozygosity and inbreeding to assess the long- and short-term population history of the Dryas
484 monkey. Pairwise Sequential Markovian Coalescent (PSMC) analysis of the Dryas monkey genome

485 revealed a dynamic evolutionary history, with a marked increase in effective population size starting
486 ca. 600,000 years ago, followed by continuous decline in the last ~150,000 years (Figure 3A). To
487 eliminate the possibility that our PSMC inferences are driven by the increased heterozygosity due to
488 gene flow, we removed all putatively introgressed regions and repeated the PSMC analysis, which
489 produced a highly similar result (Figure 3A). We therefore suggest that the population size increase in
490 the Dryas monkey and the associated likely range expansion facilitated secondary contact between the
491 Dryas monkey and the vervets.
492 As previously reported, low genomic coverage shifts the PSMC trajectory and makes inference less
493 reliable, particularly for more recent time periods (Nadachowska-Brzyska et al., 2016). Therefore, to
494 allow for demographic comparisons to the vervets, we re-ran the PSMC on the Dryas monkey genome
495 down-sampled to similar coverage as the genomic data available for the vervets. We indeed find a
496 strong shift of the PSMC trajectory as a result of reduced coverage to more recent times and lower
497 effective population size (Figure 3A). Nonetheless, bias related to coverage affect the different vervet
498 genomes in a similar manner (Nadachowska-Brzyska et al., 2016), thus relative comparison between
499 the species suggest that 100,000 – 300,000 years ago the Dryas monkey population was the largest
500 among all vervets (Figure 3B).

501 The genetic diversity of the Dryas monkey (measured as between chromosome-pair differences), a
502 proxy for the adaptive potential of a population (Lande and Shannon, 1996), is high compared to that
503 of the much more abundant vervets (Figure 4A). The Dryas monkey individual also shows no signs of
504 excessive recent inbreeding, which would manifest itself in a high fraction of the genome contained in
505 long tracts of homozygosity (> 2.5 Mb) (Figure 4B). To estimate genetic load, we identified all genes in
506 the Dryas monkey genome containing one or more loss-of-function (LoF) mutations and identified all
507 missense mutations in genes other than those already containing LoF-mutation(s) (as such mutations
508 likely behave neutrally). We find multiple genes in the Dryas monkey containing a homozygous LoF
509 mutation associated with a disease phenotype in humans (n = 27), including in SEPT12, associated with
510 reduced sperm mobility (Kuo *et al.*, 2015) and in SLAMF9, associated with reduced immunity to
511 tapeworm infections (Cárdenas *et al.*, 2014). However, genome-wide measures of genetic load,
512 measured as the ratio between LoF or missense and synonymous mutations (dN/dS), do not show an
513 increased genomic burden of deleterious mutations in the Dryas monkey compared to the much more
514 abundant and widely distributed vervets (Figure 4C-D). The demographic history and the genome-wide
515 measures of genetic diversity and genetic load of the Dryas monkey thus suggest that the population
516 of this endangered primate might be larger than currently recognised and has good chances for long-
517 term survival, if appropriate conservation measures are implemented.

518 **Acknowledgments**

We thank the DRC government and ICCN for facilitating sample collection and the 200 Mammals Consortium for providing assemblies of *Cercopithecus neglectus* and *Erythrocebus patas*. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The authors acknowledge support from the Uppsala Multidisciplinary Centre for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. This work was supported by FORMAS (2016-00835) to K.G.

Data accessibility

Whole genome sequence data generated in this study is available in the European nucleotide archive under accession number PRJEB32105. Cytochrome B sequences are available in the genbank under accession number XXX.

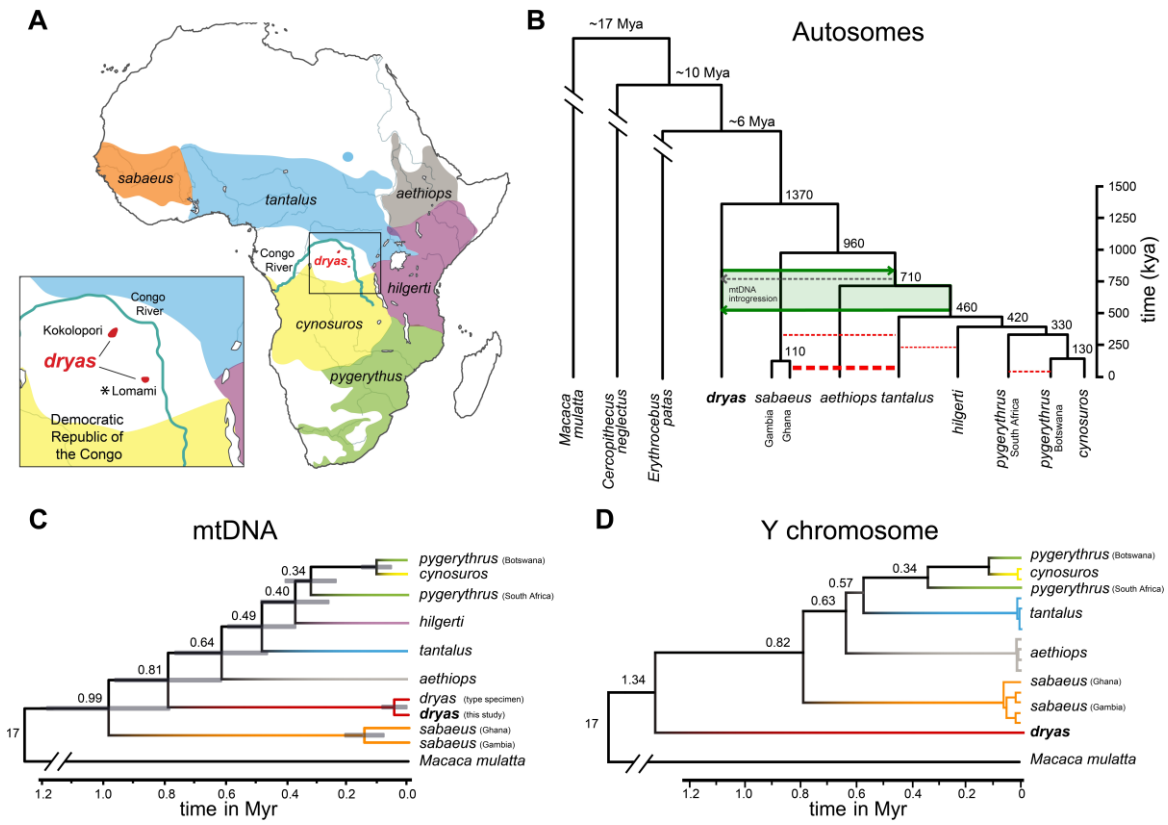
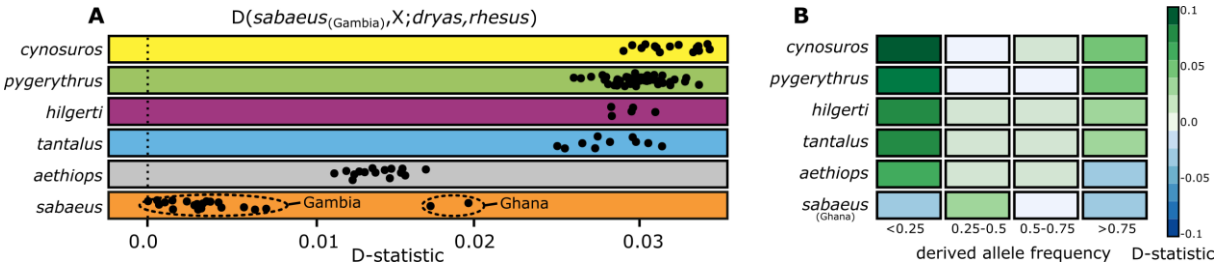


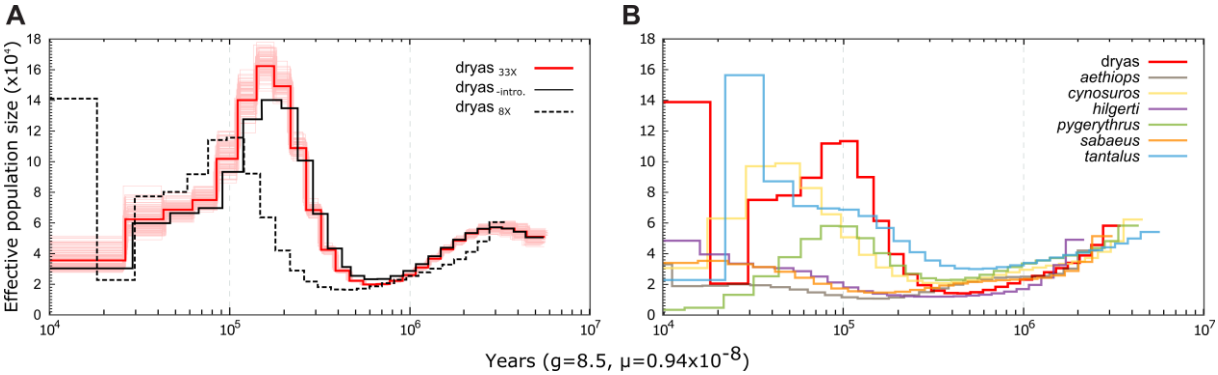
Figure 1. Phylogeny of the Dryas monkey and the vervets. (A) Distribution ranges of vervets. *Cercopithecus dryas* is currently known from only two isolated populations: The Kokolopori-Wamba and the Lomami-Lualaba interfluve. The sequenced Dryas monkey individual was sampled from the Lomami population as indicated by the asterisk. **(B)** The autosomal consensus phylogeny and MCMC dating of the Dryas monkey and the vervets in comparison to two other guenon genomes, using the rhesus macaque (*Macaca mulatta*) as outgroup. The tree topology is supported by multiple analyses

538 (see Methods, Figure S2, S3). Red dotted lines depict previously reported admixture events between
539 the different vervet species (Svardal *et al.*, 2017), with the line width corresponding to the relative
540 strength of admixture. Green arrows show the best-supported timing of admixture between the Dryas
541 monkey and the vervets. *Chl. sabaeus* individuals from Ghana have a higher proportion of shared
542 derived alleles with the Dryas monkey than *Chl. sabaeus* individuals from Gambia, which is the result
543 of a strong secondary admixture pulse between *Chl. tantalus* and the Ghana *Chl. sabaeus* population
544 (Svardal *et al.*, 2017). Numbers at the nodes represent estimated mean divergence times (see Figure
545 S3). (**C**) Mitochondrial phylogeny and divergence times, provided as numbers at the nodes. Grey blocks
546 show the 95% confidence intervals as obtained with MCMCTree. (**D**) UPGMA Y-chromosomal
547 phylogeny, calibrated using the rhesus macaque Y-chromosome as outgroup. Number at the nodes
548 represent estimated divergence times.

549 

550 **Figure 2. Genome-wide statistics support gene flow between the Dryas monkey and the vervets** (**A**)
551 Obtained D-statistics for individual comparisons. Each dot depicts a comparison of a single vervet
552 genome (X) with the *Chl. sabaeus* individual that showed the least amount of shared derived alleles
553 with the Dryas monkey in all individual comparisons. (**B**) D-statistics stratified by derived allele
554 frequency for each species, using all *Chl. sabaeus*$_{\text{Gambia}}$ individuals as ingroup.

555

556 

557 **Figure 3. Demographic history of the Dryas monkey and the vervets** (**A**) PSMC analyses for the repeat-
558 filtered Dryas monkey genome (red), after removing putatively introgressed regions (black), and after
559 down-sampling to 8x coverage (dotted line). As the curve is strongly shifted at lower coverage, the
560 down-sampled genome was used for between-species comparison in (B). (B) PSMC analysis of medium
561 coverage genomes (7.4-9.8X) for all vervets and the down-sampled Dryas monkey genome (8X).

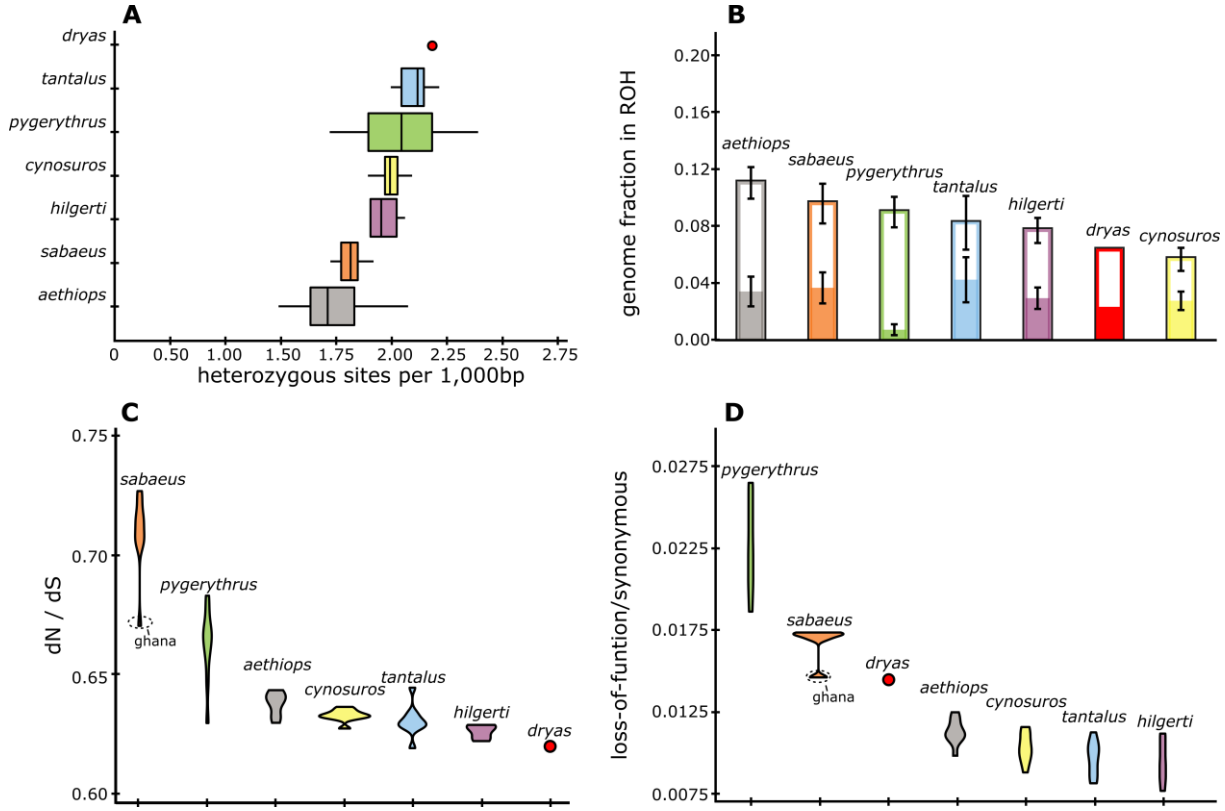


Figure 4. Genome-wide measures of genetic diversity and genetic load in the Dryas monkey (**A**) Average number of heterozygous sites per 1,000 base pairs. (**B**) Fraction of the genome in runs of homozygosity (ROH) above 100 kb (open bars) and fraction of the genome in ROH > 2.5 Mb (solid bars). (**C**) Ratio of missense to synonymous mutations, excluding all missense mutations within genes containing one or more LOF mutations. (**D**) Ratio of LOF to synonymous mutations, counting genes with more than one LOF mutation only once.

References

Van der Auwera, G. A. *et al.* (2013) 'From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline', in *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 11.10.1-11.10.33. doi: 10.1002/0471250953.bi1110s43.

Beadle, L. (1981) *The inland waters of tropical Africa: an introduction to tropical limnology*. 2nd editio. London/New York.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data.', *Bioinformatics (Oxford, England)*. Oxford University Press, 30(15), pp. 2114–20. doi: 10.1093/bioinformatics/btu170.

Butynski, T. M. (2013) 'Cercopithecus dryas', in Kalina, T. M. B. and J. (ed.) *The Mammals of Africa*. Bloomsbury Publishing, London, pp. 306–309.

Cahill, J. A. *et al.* (2016) 'Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 371(1699). doi: 10.1098/rstb.2015.0138.

Cárdenas, G. *et al.* (2014) 'Neurocysticercosis: the effectiveness of the cysticidal treatment could be influenced by the host immunity', *Medical Microbiology and Immunology*, 203(6), pp. 373–381. doi: 10.1007/s00430-014-0345-2.

Chen, H., Patterson, N. and Reich, D. (2010) 'Population differentiation as a test for selective sweeps', *Genome Research*, 20(3), pp. 393–402. doi: 10.1101/gr.100545.109.

Colyn, Gautier-Hion, A. and van den Audenaerde, T. (1991) 'Cercopithecus dryas Schwarz 1932 and C. salongo Thys van den Audenaerde 1977 are the same species with an age-related coat pattern', *Folia Primatologica*. Karger Publishers, 56(3), pp. 167–170. doi: 10.1159/000156543.

Colyn, M. (1987) 'Les primates des forêts ombrophiles de la cuvette du Zaïre: interprétations zoogéographiques des modèles de distribution', *Revue De Zoologie Africaine*, 101, pp. 183–196.

Colyn, M. and Deleporte, P. (2004) 'Biogeographic Analysis of Central African Forest Guenons', in *The Guenons: Diversity and Adaptation in African Monkeys*. Boston: Kluwer Academic Publishers, pp. 61–78. doi: 10.1007/0-306-48417-X_6.

Colyn, M., Gautier-Hion, A. and Verheyen, W. (1991) 'A Re-Appraisal of Palaeoenvironmental History in Central Africa: Evidence for a Major Fluvial Refuge in the Zaire Basin', *Journal of Biogeography*,

599 18(4), p. 403. doi: 10.2307/2845482.

600 Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*. Oxford University
601 Press, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

602 Debaisieux, S. *et al.* (2012) 'The Ins and Outs of HIV-1 Tat', *Traffic*. John Wiley & Sons, Ltd (10.1111),
603 13(3), pp. 355–363. doi: 10.1111/j.1600-0854.2011.01286.x.

604 Delaneau, O. *et al.* (2013) 'Haplotype Estimation Using Sequencing Reads', *The American Journal of*
605 *Human Genetics*, 93(4), pp. 687–696. doi: 10.1016/j.ajhg.2013.09.002.

606 DePristo, M. A. *et al.* (2011) 'A framework for variation discovery and genotyping using next-
607 generation DNA sequencing data', *Nature Genetics*. Nature Publishing Group, 43(5), pp. 491–498.
608 doi: 10.1038/ng.806.

609 Dierckxsens, N., Mardulyn, P. and Smits, G. (2016) 'NOVOPlasty: De novo assembly of organelle
610 genomes from whole genome data', *Nucleic Acids Research*. Oxford University Press, 45(4), p.
611 gkw955. doi: 10.1093/nar/gkw955.

612 Eriksson, J. *et al.* (2004) 'Rivers influence the population genetic structure of bonobos (Pan
613 paniscus)', *Molecular Ecology*, 13(11), pp. 3425–3435. doi: 10.1111/j.1365-294X.2004.02332.x.

614 Fay, J. C., Wyckoff, G. J. and Wu, C. I. (2001) 'Positive and negative selection on the human genome.',
615 *Genetics*, 158(3), pp. 1227–34.

616 Felsenstein, J. (2005) 'PHYLIP (Phylogeny Inference Package)'. Distributed by the author. Department
617 of Genome Sciences, University of Washington, Seattle.

618 Fumagalli, M. (2013) 'Assessing the effect of sequencing depth and sample size in population
619 genetics inferences', *PLoS ONE*. Edited by L. Orlando, 8(11), p. e79667. doi:
620 10.1371/journal.pone.0079667.

621 Fumagalli, M. *et al.* (2013) 'Quantifying population genetic differentiation from next-generation
622 sequencing data', *Genetics*, 195(3), pp. 979–992. doi: 10.1534/genetics.113.154740.

623 Gotz, S. *et al.* (2008) 'High-throughput functional annotation and data mining with the Blast2GO
624 suite', *Nucleic Acids Research*, 36(10), pp. 3420–3435. doi: 10.1093/nar/gkn176.

625 Green, R. E. *et al.* (2010) 'A draft sequence of the Neandertal genome.', *Science (New York, N.Y.)*,
626 328(5979), pp. 710–722. doi: 10.1126/science.1188021.

627 Grubb, P. *et al.* (2003) 'Assessment of the Diversity of African Primates', *International Journal of*

Primatology. Springer New York, 24(6), pp. 1301–1357. doi: 10.1023/B:IJOP.0000005994.86792.b9.

Guschanski, K. *et al.* (2013) 'Next-generation museomics disentangles one of the largest primate radiations', *Systematic Biology*, 62(4), pp. 539–554. doi: 10.1093/sysbio/syt018.

Hart, J. A. *et al.* (2019) 'Cercopithecus dryas. The IUCN Red List of Threatened Species'.

Haus, T. *et al.* (2013) 'Mitochondrial diversity and distribution of African green monkeys (chlorocebus gray, 1870).', *American journal of primatology*. Wiley-Blackwell, 75(4), pp. 350–60. doi: 10.1002/ajp.22113.

Hawks, J. (2017) 'Introgression Makes Waves in Inferred Histories of Effective Population Size', *Human Biology*, 89(1), pp. 67–80. doi: 10.13110/humanbiology.89.1.04.

Hubisz, M. J., Pollard, K. S. and Siepel, A. (2011) 'PHAST and RPHAST: phylogenetic analysis with space/time models', *Briefings in Bioinformatics*, 12(1), pp. 41–51. doi: 10.1093/bib/bbq072.

Huson, D. H. and Bryant, D. (2006) 'Application of Phylogenetic Networks in Evolutionary Studies', *Molecular Biology and Evolution*, 23(2), pp. 254–267. doi: 10.1093/molbev/msj030.

Juric, I., Aeschbacher, S. and Coop, G. (2016) 'The Strength of Selection against Neanderthal Introgression', *PLOS Genetics*. Edited by D. Reich. Public Library of Science, 12(11), p. e1006340. doi: 10.1371/journal.pgen.1006340.

Kennis, J. *et al.* (2011) 'The impact of the Congo River and its tributaries on the rodent genus Praomys: Speciation origin or range expansion limit?', *Zoological Journal of the Linnean Society*. Oxford University Press, 163(3), pp. 983–1002. doi: 10.1111/j.1096-3642.2011.00733.x.

Kim, H., Kang, K. and Kim, J. (2009) 'AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2.', *Nucleic acids research*, 37(9), pp. 2940–50. doi: 10.1093/nar/gkp149.

Korneliussen, T. S., Albrechtsen, A. and Nielsen, R. (2014) 'ANGSD: Analysis of Next Generation Sequencing Data', *BMC Bioinformatics*. BioMed Central, 15(1), p. 356. doi: 10.1186/s12859-014-0356-4.

Kuderna, L. F. K. *et al.* (2019) 'Selective single molecule sequencing and assembly of a human Y chromosome of African origin', *Nature Communications*. Nature Publishing Group, 10(1), p. 4. doi: 10.1038/s41467-018-07885-5.

Kuhlwilm, M. *et al.* (2016) 'Ancient gene flow from early modern humans into Eastern Neanderthals', *Nature*. Nature Publishing Group, 530(7591), pp. 429–433. doi: 10.1038/nature16544.

Kumar, S. *et al.* (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Molecular Biology and Evolution*. Edited by F. U. Battistuzzi. Oxford University Press, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kuo, Y.-C. *et al.* (2015) 'SEPT12 orchestrates the formation of mammalian sperm annulus by organizing core octameric complexes with other SEPT proteins', *Journal of Cell Science*, 128(5), pp. 923–934. doi: 10.1242/jcs.158998.

Kuroda, S., Kano, T. and Muhindo, K. (1985) 'Further information on the new monkey species,Cercopithecus salongo Thys van den Audenaerde, 1977', *Primates*. Springer-Verlag, 26(3), pp. 325–333. doi: 10.1007/BF02382406.

Lande, R. and Shannon, S. (1996) 'The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment', *Evolution*, 50(1), p. 434. doi: 10.2307/2410812.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM'. doi: arXiv:1303.3997.

Li, H. and Durbin, R. (2011) 'Inference of human population history from individual whole-genome sequences', *Nature*. Nature Research, 475(7357), pp. 493–496. doi: 10.1038/nature10231.

Liang, M. and Nielsen, R. (2014) 'The Lengths of Admixture Tracts', *Genetics*, 197(3), pp. 953–967. doi: 10.1534/genetics.114.162362.

de Manuel, M. *et al.* (2016) 'Chimpanzee genomic diversity reveals ancient admixture with bonobos', *Science*, 354(6311).

Martin, S. H. and Van Belleghem, S. M. (2017) 'Exploring Evolutionary Relationships Across the Genome Using Topology Weighting.', *Genetics*. Genetics, 206(1), pp. 429–438. doi: 10.1534/genetics.116.194720.

Martin, S. H., Davey, J. W. and Jiggins, C. D. (2015) 'Evaluating the use of ABBA-BABA statistics to locate introgressed loci', *Molecular Biology and Evolution*. Oxford University Press, 32(1), pp. 244–257. doi: 10.1093/molbev/msu269.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.', *Genome research*. Cold Spring Harbor Laboratory Press, 20(9), pp. 1297–303. doi: 10.1101/gr.107524.110.

McLaren, W. *et al.* (2016) 'The Ensembl Variant Effect Predictor', *Genome Biology*. BioMed Central, 17(1), p. 122. doi: 10.1186/s13059-016-0974-4.

Nadachowska-Brzyska, K. *et al.* (2016) 'PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers.', *Molecular ecology*. Wiley-Blackwell, 25(5), pp. 1058–72. doi: 10.1111/mec.13540.

Paradis, E., Claude, J. and Strimmer, K. (2004) 'APE: Analyses of Phylogenetics and Evolution in R language', *Bioinformatics*. Oxford University Press, 20(2), pp. 289–290. doi: 10.1093/bioinformatics/btg412.

Pfeifer, S. P. (2017) 'Direct estimate of the spontaneous germ line mutation rate in African green monkeys', *Evolution*. John Wiley & Sons, Ltd (10.1111), 71(12), pp. 2858–2870. doi: 10.1111/evo.13383.

Pickrell, J. K. and Pritchard, J. K. (2012) 'Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data', *PLoS Genetics*. Edited by H. Tang. Public Library of Science, 8(11), p. e1002967. doi: 10.1371/journal.pgen.1002967.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS ONE*. Edited by A. F. Y. Poon. Public Library of Science, 5(3), p. e9490. doi: 10.1371/journal.pone.0009490.

Purcell, S. *et al.* (2007) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Reis, M. d. and Yang, Z. (2011) 'Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times', *Molecular Biology and Evolution*, 28(7), pp. 2161–2172. doi: 10.1093/molbev/msr045.

Reis, M. Dos *et al.* (2018) 'Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case', *Systematic Biology*. Edited by S. Ho. Narnia, 67(4), pp. 594–615. doi: 10.1093/sysbio/syy001.

Sankararaman, S. *et al.* (2014) 'The genomic landscape of Neanderthal ancestry in present-day humans', *Nature*. Nature Publishing Group, 507(7492), pp. 354–357. doi: 10.1038/nature12961.

Sayyari, E. and Mirarab, S. (2016) 'Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies', *Molecular Biology and Evolution*, 33(7), pp. 1654–1668. doi: 10.1093/molbev/msw079.

718 Schliep, K. P. (2011) ‘phangorn: phylogenetic analysis in R’, *Bioinformatics*. Oxford University Press,
719 27(4), pp. 592–593. doi: 10.1093/bioinformatics/btq706.

720 Schumer, M. *et al.* (2018) ‘Natural selection interacts with recombination to shape the evolution of
721 hybrid genomes.’, *Science (New York, N.Y.)*. NIH Public Access, 360(6389), pp. 656–660. doi:
722 10.1126/science.aar3684.

723 Schwarz, E. (1932) ‘Der Vertreter der Diana-Meerkatze in Zentral-Afrika’, *Rev Zoologique Botanique*
724 *Africaine*, (21), pp. 251–254.

725 von Schwedler, U. K. *et al.* (2003) ‘The protein network of HIV budding.’, *Cell*, 114(6), pp. 701–13.

726 Sievers, F. *et al.* (2011) ‘Fast, scalable generation of high-quality protein multiple sequence
727 alignments using Clustal Omega.’, *Molecular systems biology*, 7(1), p. 539. doi: 10.1038/msb.2011.75.

728 Slatkin, M. and Pollack, J. L. (2008) ‘Subdivision in an ancestral species creates asymmetry in gene
729 trees.’, *Molecular biology and evolution*. Oxford University Press, 25(10), pp. 2241–6. doi:
730 10.1093/molbev/msn172.

731 Stankiewicz, J. and de Wit, M. J. (2006) ‘A proposed drainage evolution model for Central Africa—Did
732 the Congo flow east?’, *Journal of African Earth Sciences*. Pergamon, 44(1), pp. 75–84. doi:
733 10.1016/J.JAFREARSCI.2005.11.008.

734 Stanton, D. W. G. *et al.* (2014) ‘Distinct and Diverse: Range-Wide Phylogeography Reveals Ancient
735 Lineages and High Genetic Variation in the Endangered Okapi (Okapia johnstoni)’, *PLoS ONE*. Edited
736 by V. Ketmaier. Public Library of Science, 9(7), p. e101081. doi: 10.1371/journal.pone.0101081.

737 Strack, B. *et al.* (2003) ‘AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus
738 budding.’, *Cell*, 114(6), pp. 689–99.

739 Sturm, R. A. *et al.* (1993) ‘Chromosomal Structure and Expression of the Human OTF1 Locus Encoding
740 the Oct-1 Protein’, *Genomics*, 16(2), pp. 333–341. doi: 10.1006/geno.1993.1194.

741 Svardal, H. *et al.* (2017) ‘Ancient hybridization and strong adaptation to viruses across African vervet
742 monkey populations’, *Nature Genetics*. Nature Publishing Group, 49(12), pp. 1705–1713. doi:
743 10.1038/ng.3980.

744 Tamura, K. and Nei, M. (1993) ‘Estimation of the number of nucleotide substitutions in the control
745 region of mitochondrial DNA in humans and chimpanzees.’, *Molecular Biology and Evolution*, 10(3),
746 pp. 512–26. doi: 10.1093/oxfordjournals.molbev.a040023.

747 Tomaszkiewicz, M., Medvedev, P. and Makova, K. D. (2017) ‘Y and W Chromosome Assemblies:

748 Approaches and Discoveries', *Trends in Genetics*, 33(4), pp. 266–282. doi: 10.1016/j.tig.2017.01.008.

749 van der Valk, T. *et al.* (2019) 'Historical Genomes Reveal the Genomic Consequences of Recent
750 Population Decline in Eastern Gorillas', *Current Biology*, 29(1), pp. 165-170.e6. doi:
751 10.1016/j.cub.2018.11.055.

752 Warren, W. C. *et al.* (2015) 'The genome of the vervet (Chlorocebus æthiops sabæus)', *Genome*
753 *Research*, 25(12), pp. 1921–1933. doi: 10.1101/gr.192922.115.

754 Xue, C. *et al.* (2016) 'The population genomics of rhesus macaques (Macaca mulatta) based on
755 whole-genome sequences.', *Genome research*. Cold Spring Harbor Laboratory Press, 26(12), pp.
756 1651–1662. doi: 10.1101/gr.204255.116.

757 Xue, Y. *et al.* (2015) 'Mountain gorilla genomes reveal the impact of long-term population decline
758 and inbreeding.', *Science (New York, N.Y.)*. NIH Public Access, 348(6231), pp. 242–5. doi:
759 10.1126/science.aaa3952.

760 Yang, Z. (2007) 'PAML 4: Phylogenetic Analysis by Maximum Likelihood', *Molecular Biology and*
761 *Evolution*, 24(8), pp. 1586–1591. doi: 10.1093/molbev/msm088.

762 Zhang, C. *et al.* (2018) 'ASTRAL-III: polynomial time species tree reconstruction from partially
763 resolved gene trees', *BMC Bioinformatics*. BioMed Central, 19(S6), p. 153. doi: 10.1186/s12859-018-
764 2129-y.

765 Zhou, C. and Rana, T. M. (2002) 'A bimolecular mechanism of HIV-1 Tat protein interaction with RNA
766 polymerase II transcription elongation complexes.', *Journal of molecular biology*, 320(5), pp. 925–42.

767 Zinner, D. *et al.* (2013) *Family Cercopithecidae (Old World Monkeys). In Handbook of the Mammals of*
768 *the World: Primates (pp. 550–753). Lynx Edicions.*

769

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
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. 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		

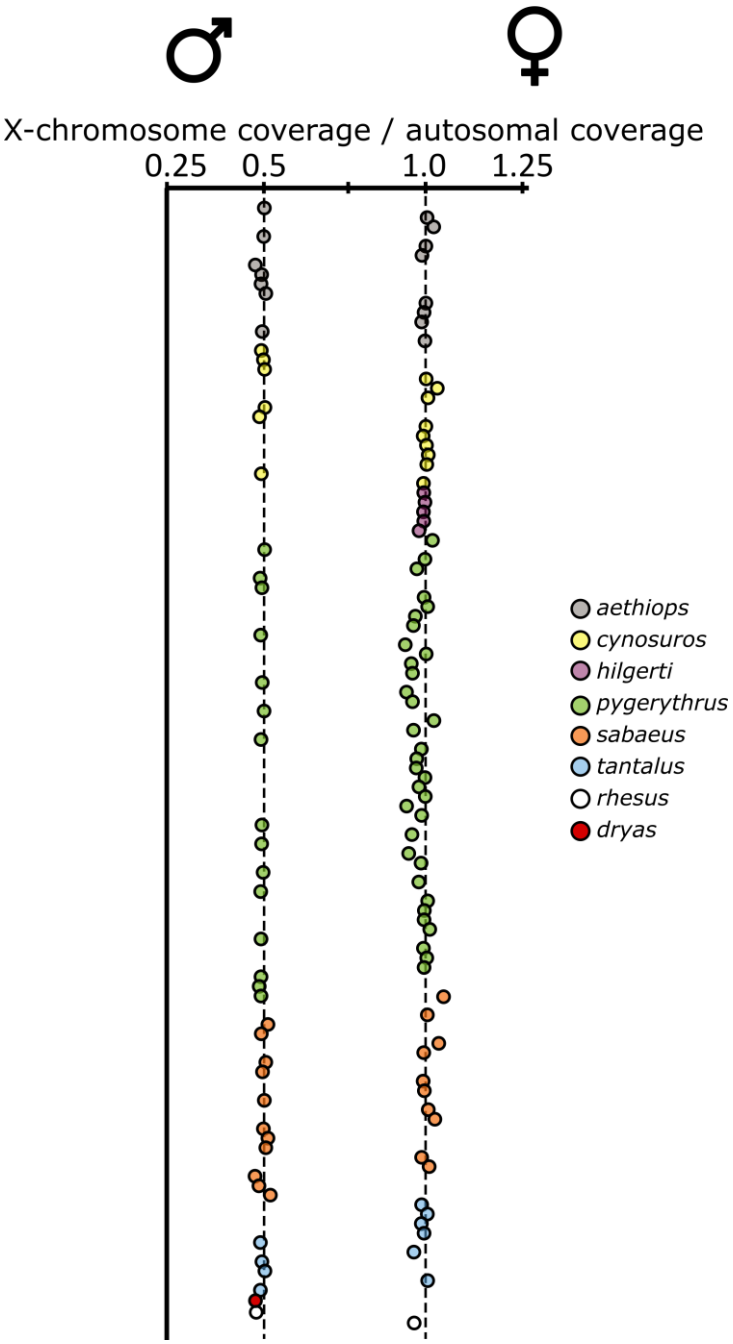


Figure S1. Sexing individuals based on the ratio of X-chromosomal to autosomal coverage.

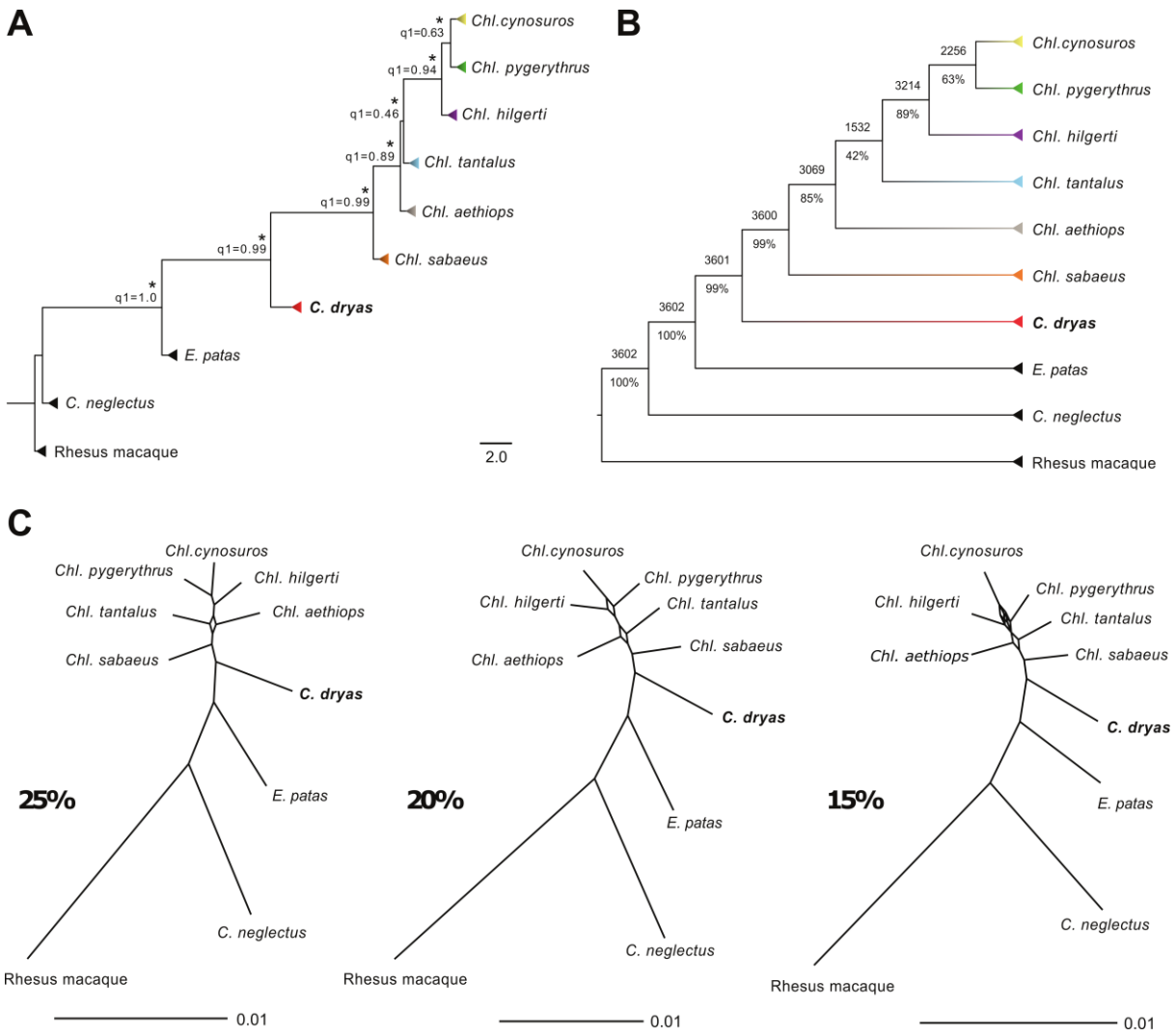


Figure S2. Phylogeny of the Dryas monkey and the vervets. (A) MSC-based species trees generated by ASTRAL using 3602 autosomal genomic windows. The tree was rooted with the rhesus macaque. Branch lengths are given in coalescent units and are an indicator of gene-tree discordance. The normalized quartet 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%.

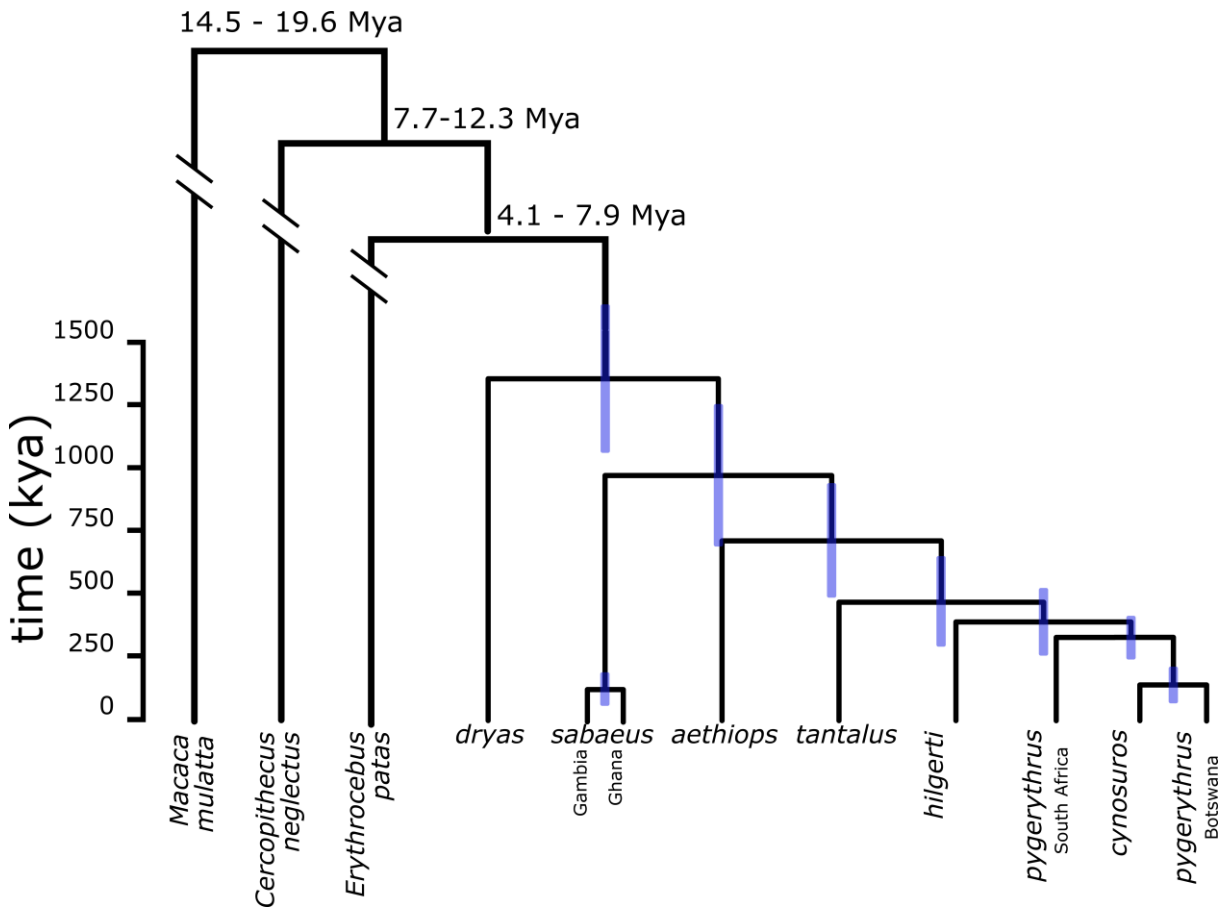


Figure S3. The autosomal consensus phylogeny and MCMC dating. Same as figure 1B but depicting 95% confidence intervals for the estimated divergence times. Confidence intervals for the oldest nodes are depicted with numbers for clarity.

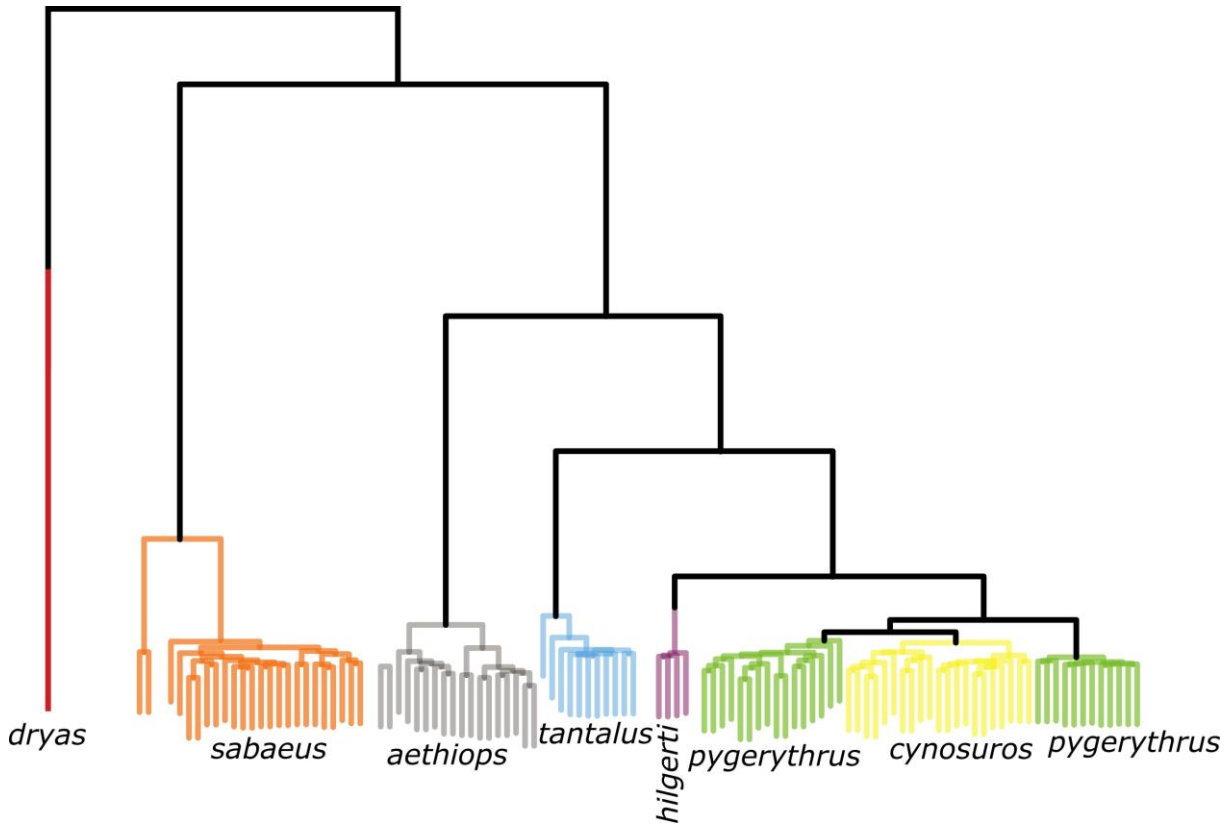


Figure S4. Evolutionary relationships of the Dryas monkey and the vervets. The tree shows the UPGMA clustering of all individuals used in this study based on all bi-allelic autosomal sites.

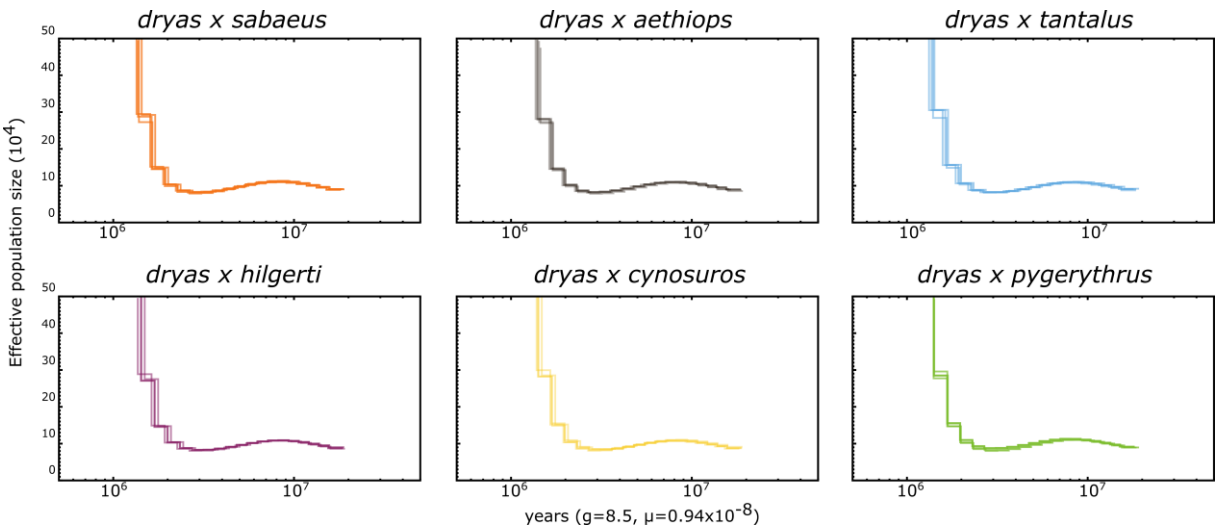


Figure S5. Artificial hybrid PSMC (hPSMC) of the Dryas monkey and the vervets. hPSMC was run on pairwise haploidized genomes for the Dryas monkey and four individuals from each vervet species assuming a generation time of 8.5 years and a mutation rate per generation of 0.94×10^{-8}. PSMC plots on artificial 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.

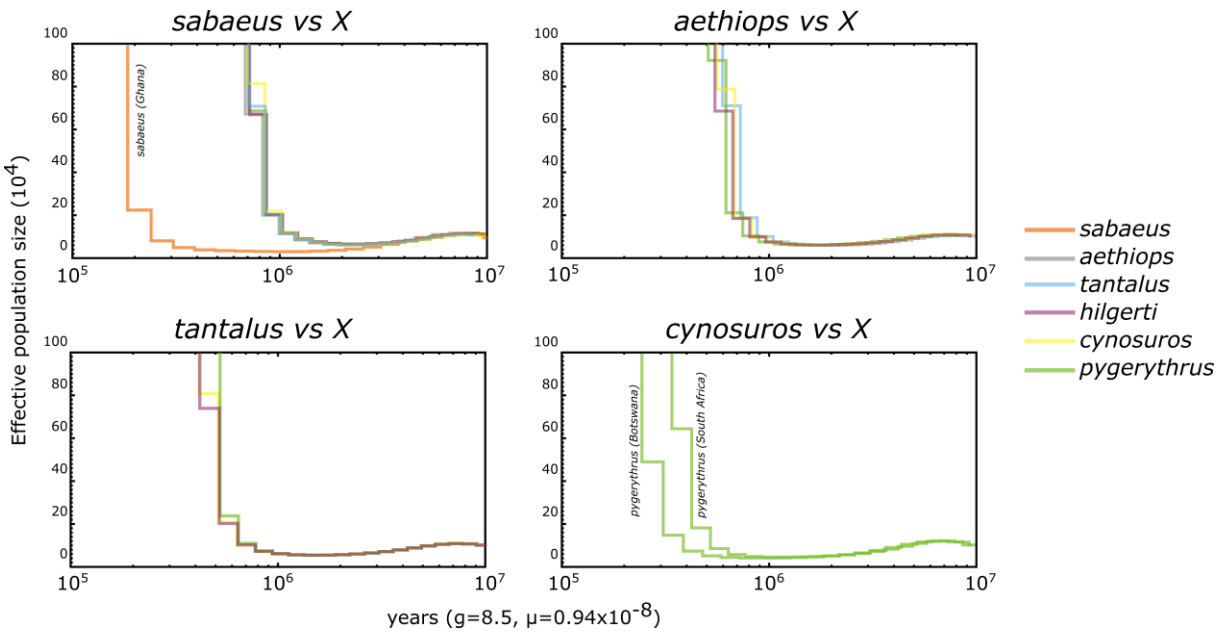


Figure S6. Artificial hybrid PSMC (hPSMC) within the vervet clade. Same as Figure S4, estimating divergence-times within the vervet genus.

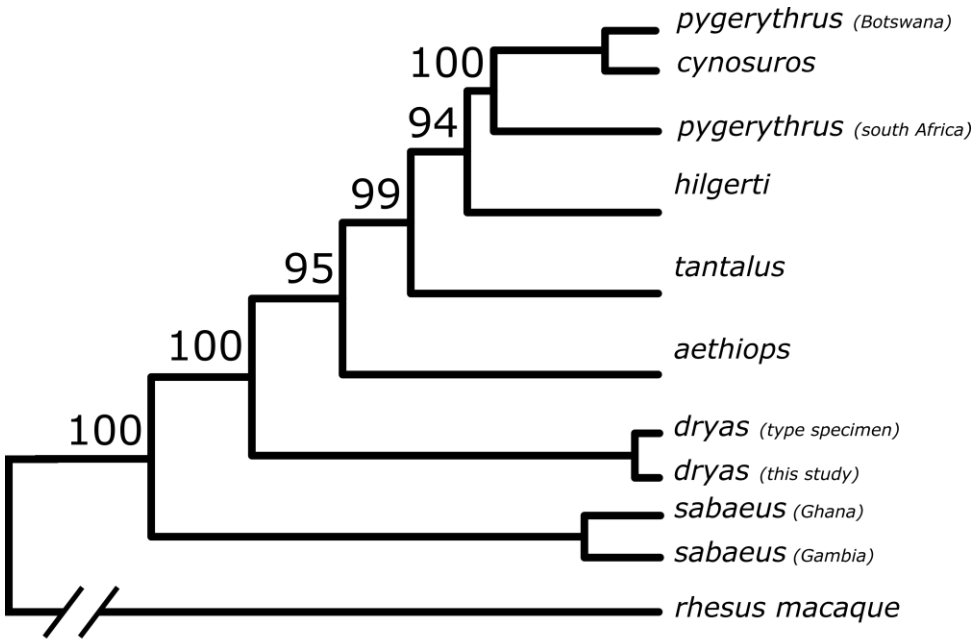


Figure S7. Maximum likelihood tree based on complete mitochondrial genomes for the vervets and the Dryas monkey (same as Figure 1C). Numbers depict bootstrap support values (1000 bootstrap replicates were run).

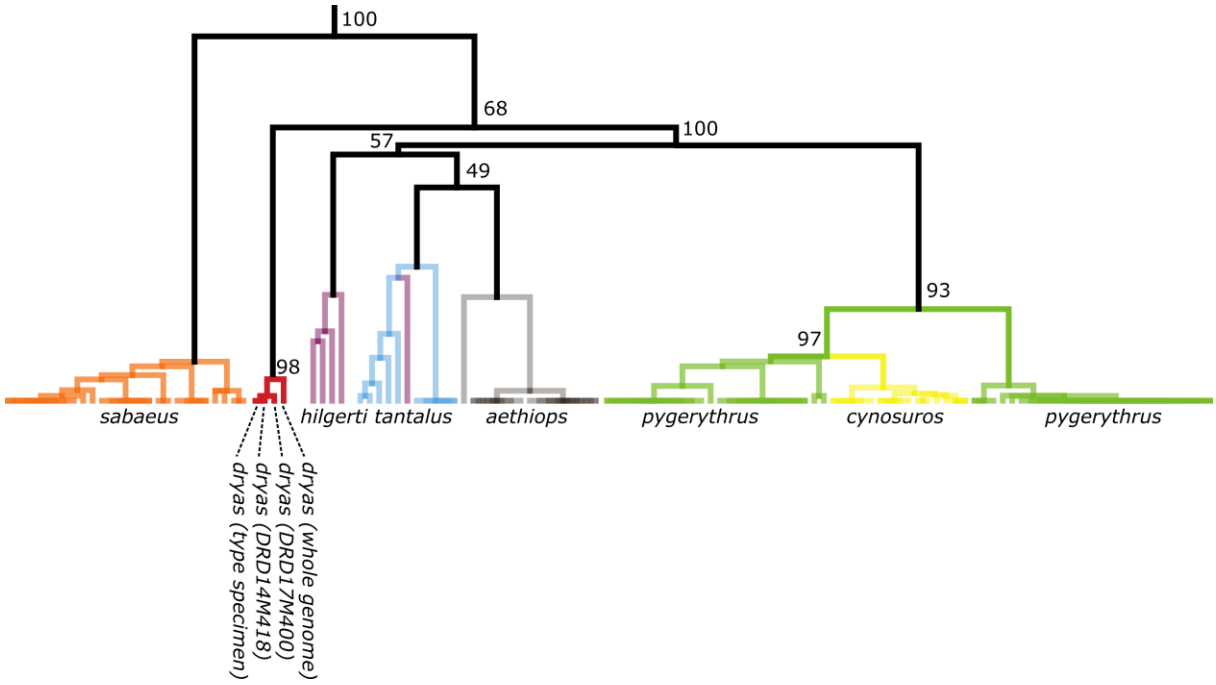


Figure S8. Maximum likelihood tree based on the cytochrome B sequence for the vervets and the Dryas monkey. Numbers depict bootstrap support values (1000 bootstrap replicates were run). Note the close clustering of all available Dryas monkey cytochrome B sequences.

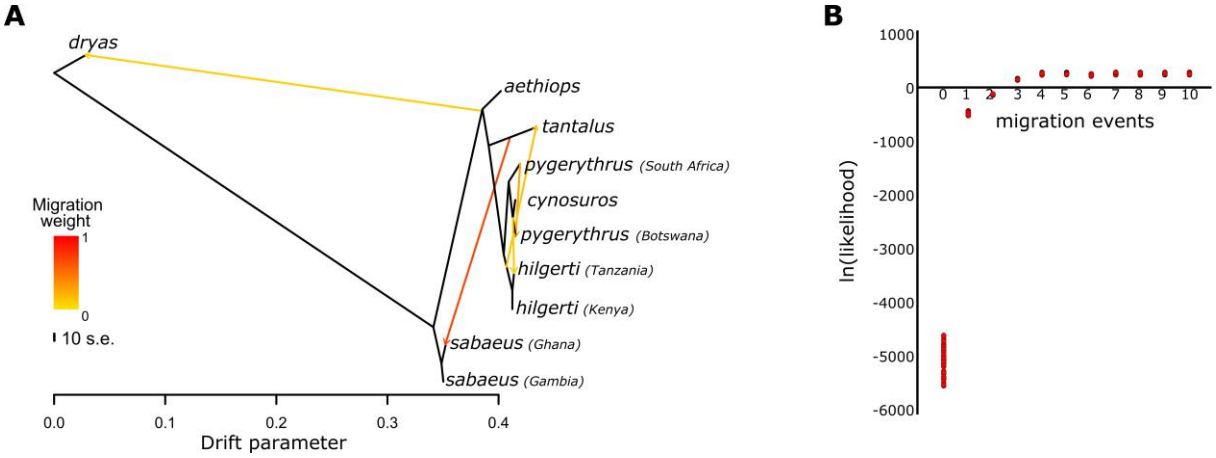


Figure S9. Model-based estimates of gene-flow. (A) TreeMix analysis for five migration events (m=5). Population allele frequencies are separated by species and geographic origin. **(B)** Likelihood support values for TreeMix models with 0 to 10 migration events respectively. After modelling five migration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

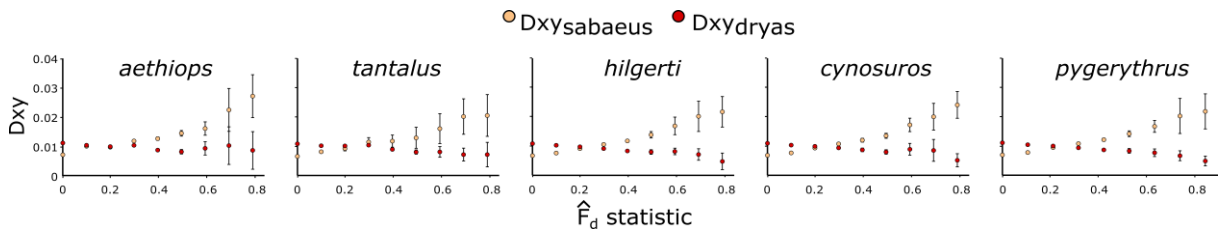


Figure S10. Identifying introgressed regions with Dxy statistic stratified by $\hat{F}_d$. Errors bars show ±3SE. Window size = 10kb. Windows with high $\hat{F}_d$ statistic (the excess of shared derived alleles with the Dryas monkey over that with *Chl. sabaeus*) have on average unusual large genetic distance (D_{xy}) to *Chl. sabaeus* and unusual low genetic distance to the Dryas monkey. Note that the vast majority of windows is at $\hat{F}_d$ ~ 0 (>90%). $D_{xy\text{-}sabaeus}$ at windows with $\hat{F}_d$ = 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 $\hat{F}_d$, $D_{xy\text{-}dryas}$ is around ½-⅓ that of $\hat{F}_d$ = 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).

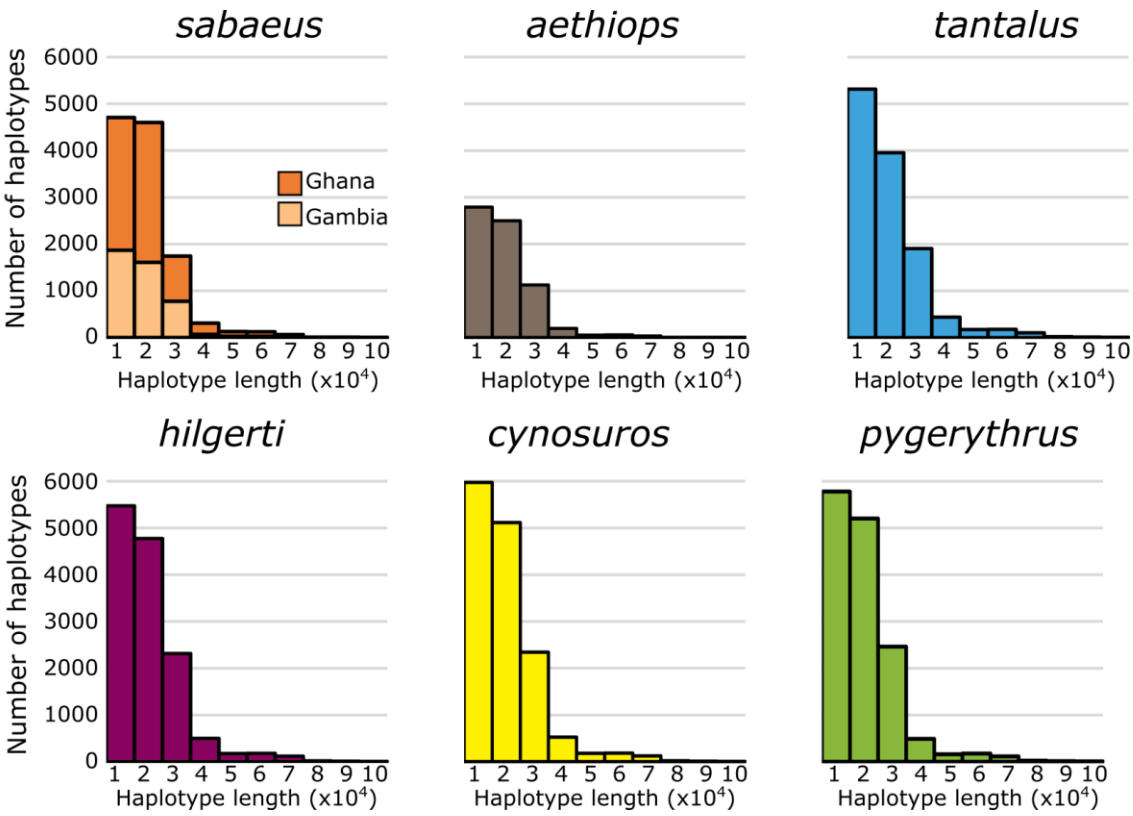


Figure S11. Length distribution of putatively introgressed haplotypes. Windows are short on average for all species. *Chl. sabaeus* individuals from Gambia have the lowest number of putatively introgressed windows, which were likely introduced into this species via secondary gene-flow with *Chl. tantalus*.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

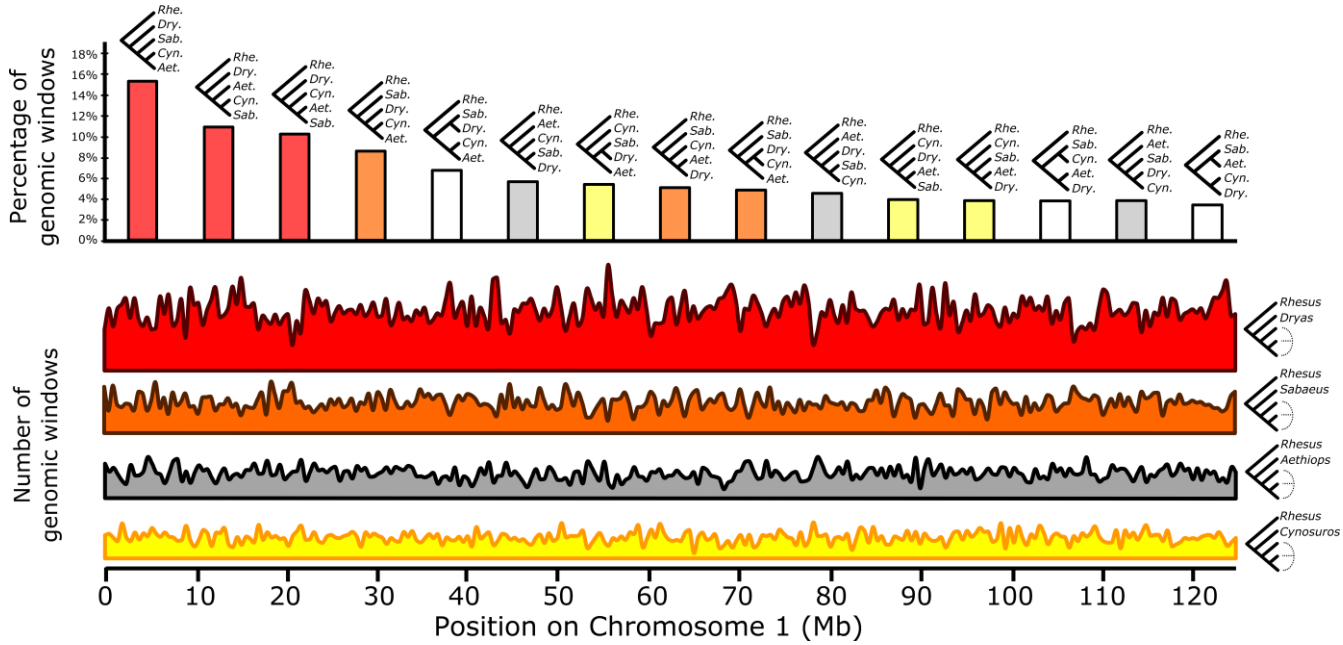


Figure S12. Exploring evolutionary relationships across the genome using topology weighting (*Twisst*). The top graph shows how often a depicted topology is supported among all non-overlapping 50 SNP windows (Rhe. = Rhesus macaque, Dry. = Dryas monkey, Sab. = *Chl. sabaeus*, Cyn. = *Chl. cynosuros*, 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).