# Supporting Information

## 1. Supplementary Methods



## 2. Supplementary Tables



## 3. Supplementary Figures

# Supplementary Methods

### 1.1. *Genotype calling from RRBS data*

We used BisSNP to call genotypes from the RRBS data using default parameters (Liu et al. 2012). To minimize genotyping error, we retained only biallelic variable sites identified in an independently collected whole genome resequencing data set (the Baboon Genome Consortium's Diversity Panel: Rogers et al. *in press*). We also filtered for variants with a Phred-scaled variant quality score ≥ 30 and a minor allele frequency > 5% within baboons. Among these variants, we further filtered for genotype calls with minimum read coverage > 4 and removed variants where genotype calls were retained for fewer than 66% of baboon individuals. This procedure resulted in a genotype call set for 49,607 biallelic sites, which together recapitulate the known baboon phylogeny (Fig. S8).

### 1.2. *DNA methylation, regions of alternate phylogeny, and nucleotide diversity*

Because incomplete lineage sorting and admixture have shaped sequence diversity in baboons (Tung and Barreiro 2017, Rogers et al. *in press*), we tested whether they also influence the relationship between genetic structure and DNA methylation levels. Using the 15 individuals sampled in the baboon diversity panel generated by the Baboon Genome Sequencing Consortium (Rogers et al. *in press*), we tested sliding 2 Mb windows (with a step size of 50 Kb) for their fit to the global consensus phylogeny shown in Fig. 1A. "Local" phylogenetic trees were generated in R using the packages *SNPrelate* (Zheng 2012) and *ape* (Paradis et al. 2004) to import genotype calls, convert genotype calls into a distance matrix, and infer a neighbor-joining tree from each window. The topology of the global phylogenetic tree was compared to the local phylogenetic tree using the all.equal function in *ape* with use.edge.lengths set to false.

On average, each window included 30,184 variable sites (note that a set of 4,000 variable sites chosen at random from the Baboon Genome Diversity Panel genotype calls perfectly recapitulated the global phylogeny in 96% of cases). Across the genome, 39% of windows perfectly matched the global consensus phylogeny (for another 33% of windows, only 1-2 individuals were misplaced relative to the consensus phylogeny). We then subset CpG sites into those that occurred in regions that perfectly matched the global phylogeny (n=542,509 CpG sites) and those that did not (n=211,852 CpG sites) and compared the covariance matrix for each set of methylation profiles to the overall genetic covariance matrix (derived from the RRBS data).

We also used the baboon diversity panel to estimate nucleotide diversity ($\pi$: Nei and Li 1979) for each baboon species. For each CpG site we profiled and each species, we calculated $\pi$ per nucleotide for the 1 kb region centered on the site. We then calculated the average value of $\pi$ across species for each CpG site, and then compared CpG sites with evidence of positive selection to those with taxonomic structure, but no evidence of positive selection. For sites with evidence for positive selection (on any lineage), we also compared mean $\pi$ between species putatively affected by the shift in selective optimum to the mean for the species that were not affected by the shift.

Finally, we used the Baboon Genome Diversity Panel to estimate how often CpG motifs near each focal CpG site (within a 1 kb window centered on the site) were disrupted among baboons. Here, we counted the number of variable sites in which one version of the site carried

an intact CpG site but the alternative version led to loss of the CpG site in at least one sequenced individual. We then used logistic regression to ask whether sites that exhibited taxonomic structure (n=20,360) were associated with higher numbers of nearby disrupted sites than sites that exhibited no taxonomic structure (n=736,002). We also asked whether sites with evidence for positive selection (on an individual baboon species using the heuristic method or on a baboon clade using both the heuristic method or OU approach; n = 2,980) were associated with higher numbers of nearby disrupted sites than sites with no such evidence (n = 17,380), in this case considering only the set of taxonomically structured sites (n=20,360).

### 1.3. *Identification of differentially methylated regions*

To calibrate the false discovery rate for sites with significant taxonomic structure (identified using ANOVA), we used the q-value approach of Storey and Tibshirani (2003) relative to an empirically determined, permutation-based null. To generate the null distributions for this analysis, we held the number of individuals sampled per species constant and conducted a series of ten permutations (at each level) where (i) individuals were randomly assigned to genus, (ii) individuals were randomly assigned to baboon clades (northern or southern) within genus, and (iii) individuals were randomly a species identity within clades. We then applied an ANOVA to generate the null distributions for differences in DNA methylation for genus-, clade-, and species-level taxonomic structure, respectively. We also used the permutation results to calculate q-values for the results of the binomial mixed effect models used test for species-specific methylation (see main text).

Following previous studies (e.g., Lister et al. 2009, Hansen et al. 2012, Lea et al. 2016), we defined differentially methylated regions (DMRs) by grouping together nearby CpG sites with similar interspecific differences in DNA methylation. To develop a null expectation, we permuted the species assignment for each sample while maintaining the structure of other covariates (e.g., bisulfite conversion rate). Across ten independent permutations, we observed a mean of 197 ± 194 s.d. CpG sites that exhibited significant clade or species-specific methylation (compared to 13,098 in the observed data). We never observed stretches of 2 kb contiguous sequence that contained 3 or more differentially methylated sites (and observed no more than 12 such stretches that contained 2 differentially methylated sites). We therefore conservatively defined DMRs as 2 kb contiguous regions that contained 3 or more differentially methylated sites.

We also checked whether these DMRs were more common than expected by chance, given the number of sites we detected that were taxonomically structured (n=20,360). Among the sites that exhibited significant taxonomic structure, we randomly assigned a label of species- or clade-specific methylation to the same number of sites detected in the real data (repeating this approach 100 times). We observed substantially more DMRs in the empirical data than expected based on our random assignment procedure (724 relative to 430.9 ± 21.4 s.d. in the label-permuted data; z-score = 13.70, $p < 10^{-6}$). Similarly, we also detected many more DMRs associated with a signature of positive selection at the species or clade levels than expected by chance (species: 5.39 ± 0.59 s.d. compared to 70 observed in the real data, z = 109.51, $p < 10^{-6}$; clade: 0.75 ± 2.09 s.d. compared to 25 observed in the real data, z = 11.60, $p < 10^{-6}$).

### 1.4. *Ornstein-Uhlenbeck model simulations and power to identify selected sites*

To estimate our power to identify cases of positive selection using the OU model approach, we simulated data consistent with genetic drift, stabilizing selection, and positive selection on either each individual baboon species or a multi-species lineage. For each site, we first simulated the mean methylation level for each species by drawing from a multivariate normal distribution (Equation 4 in the main text). We then simulated the methylation level for each individual based on the species' mean methylation levels; methylation for each individual was drawn from a normal distribution where the expectation is the mean methylation of the species to which the individual belonged and variance is $\tau^2$, the intraspecific variance. Simulated parameter values for $\sigma^2$ and $\tau^2$ came from a grid of 5 values each that spanned the range of observed values ($\sigma^2$ = 0-10; $\tau^2$ = 0-20). $\alpha$ was set to 0 to simulate genetic drift and randomly drawn from a uniform distribution between 5 and 10 for sites simulated under stabilizing or positive selection. These values for $\alpha$ correspond to strong selection, and are near the upper end of the observed parameter estimates; they thus almost completely remove phylogenetic structure from the covariance matrix. For positive selection, we simulated a 50% difference in DNA methylation level between the ancestral optimum level ($\theta_1$) and the new optimum level for the lineage on which a shift occurred ($\theta_2$). Following the approach we used for the observed data, we then filtered the complete set of simulated sites for those that exhibited interspecific differences in DNA methylation (ANOVA, FDR < 0.10). For the filtered set, we identified the best fitting Ornstein-Uhlenbeck model for each site using AIC.

Initially, we fit eight models that included positive selection on some part of the baboon tree: one each for positive selection on the five individual baboon species, one for the anubis-Guinea lineage, and one for each clade (northern and southern). Across 10,000 simulations, we calculated the proportion of cases in which sites were assigned to each model, for sites that were simulated under drift, stabilizing selection, or positive selection. For sites in which multiple best models included positive selection somewhere on the tree, we were unable to localize positive selection to a specific branch. We therefore grouped all such sites into a single, inferred "positive selection" category. Using all eight models that included positive selection (plus models for drift across the tree and stabilizing selection across the tree), we identified 63-78% of true simulated cases of positive selection. However, we also assigned 44% and 49% of sites simulated under genetic drift and stabilizing selection to positive selection, respectively.

Because of these high false positive rates, we chose to only test for positive selection on multi-species lineages (retaining three models that included positive selection: one for each clade and one for selection on the anubis-Guinea lineage). Doing so substantially reduced the false positive rate (19% for genetic drift; 17% for stabilizing selection), but still allowed us to detect 55-61% of true cases of positive selection (Table S5). As cautioned in the main text, because this approach does not achieve perfect specificity and sensitivity, the sets of sites identified in the observed data should be treated as enriched for cases of positive selection as opposed to as a definitive set. We focused our analyses on sites inferred to evolve under positive selection because stabilizing selection was rarely chosen as the best fitting model, and most true simulated cases of stabilizing selection were misassigned to genetic drift. Further, strong stabilizing selection generates little detectable taxonomic structure in the DNA methylation data (sites with no evident taxonomic structure were not considered in our tests for selection).

We also examined whether sample size and/or independent evolutionary history contribute to our power to detect lineage-specific shifts. Neither sample size nor independent evolutionary history has a large effect on the false positive rate. However, more independent evolution appears to confer increased power to detect true positive cases of positive selection.

For example, hamadryas baboons have the longest independent evolutionary history and we were better powered to detect positive selection on the hamadryas lineage than on any other individual baboon species.

### 1.5. *Correcting for heterospecific mapping biases*

Reads from all the species in our sample were mapped to the anubis baboon reference genome (*Panu2.0*), which could introduce biases due to heterospecific mapping. To assess the impact of heterospecific mapping, we generated haploid genomes and simulated RRBS reads for individuals of each baboon species. The haploid genomes were generated from the baboon genome diversity panel genotype calls based on one randomly selected individual per species. If the individual was heterozygous at a SNP, the non-reference allele was chosen to maximize the possibility of potential biases. Haploid genomes underwent *in silico Msp1* digestion using the SimRAD package in R (Lepais & Weir 2016) and fragments between 150 and 450 bp were retained, as in our actual RRBS protocol. We simulated 100 bp reads from the start and end of each fragment, with a bisulfite conversion rate of 100% assuming no non-CpG methylation (i.e., all non-CpG cytosines were *in silico* converted to uracil/thymine).

Each set of reads was duplicated to produce one data set that simulated post-bisulfite conversion reads carrying only methylated CpG sites (no change to the original sequence at CpG sites, but all other C's converted to T's), and a second data set that simulated post-bisulfite conversion reads carrying only unmethylated CpGs (all C's in the read converted to T's), but that was otherwise identical to the first. When combined, the two data sets simulate a true CpG methylation level of 0.5 at all sites. We then mapped the combined data set to *Panu2.0* and estimated CpG methylation levels after mapping, following the same procedure used for our real data.

We identified two cases in which mapping to *Panu2.0* leads to systematic inaccurate methylation level estimates (i.e., estimates not equal to 0.5) (Figure S5). The first case is when a CpG site is eliminated, relative to the reference genome, by a transition from cytosine to thymine (i.e., the site is CG in the reference genome, but TG in the sample). In this case, reads from TG individuals are treated as if they were converted to thymine and are counted as unmethylated, when in fact DNA methylation is no longer possible at those sites. This error could create a bias towards more (apparent) completely unmethylated sites with increasing genetic distance from the reference genome.

The second case is when a new CpG site is introduced in the sample genome relative to the reference genome, where the reference sequence is TG. The methylation level at the new site influences the read's mapping efficiency because unmethylated CpGs are converted to TG and match the reference sequence better than methylated CpG sites, which are retained as CG. Because CpG site methylation is only called at CpG sites that occur in the reference genome, this difference in mapping efficiency will only affect CpG methylation estimates at CpG sites that are physically close to the location of the new, non-reference site. This case can lead to incorrect DNA methylation estimates at the CpG site in the reference genome, with error rates potentially increasing with increasing genetic distance from the reference genome (the more distant the sample, the more likely 1 more mismatch will cause the read to fail to map entirely).

These problems can be resolved by (i) excluding methylation level estimates for samples in which a CpG site is disrupted by a C → T transition; and (ii) using three-nucleotide mapping, which allows Cs and Ts to map interchangeably (i.e., a cytosine base in a read can map to a thymine reference without being considered a mismatch, and vice-versa). To remove C → T

transitions, we used information from the mapped reads. For any CpG site, bisulfite sequencing reads can map representing the Watson strand (CG for methylated sites, TG for unmethylated sites), the Crick strand (GC for methylated sites, GT for unmethylated sites), the reverse Watson strand (as a result of PCR after bisulfite conversion: GC for methylated sites, AC for unmethylated sites), and the reverse Crick strand (also the result of PCR: CG for methylated sites, CA for unmethylated sites). To identify C → T transitions in the original sample, we use information from reads mapped to the Watson and Crick strands only, ignoring those mapped to the reverse Watson and reverse Crick strands. Unlike true CpG sites, where no adenines are ever detected, CpG sites that have undergone a C → T transition always map to the Watson strand as TG and to the Crick strand as AT after bisulfite conversion (if the transition is G → A at the second base, which disrupts the Crick strand CpG motif, we would observe TA reads mapping to the Watson strand and GT reads mapping to the Crick strand). Thus, observing any adenine base mapping to the reference genome CpG motif indicates disruption to the original CpG site. We therefore removed all sample-site combinations from the data set when this condition was observed.

After making these corrections, we observe no systematic errors in methylation level estimation that can be explained by genetic distance from the reference genome (Figure S7).

**1.6.** ***Assessment of the effects of cell type heterogeneity and source population***

Estimates of DNA methylation levels in heterogeneous tissues like whole blood can be affected by variation in cell type composition across samples (Houseman et al. 2012, Reinius et al. 2012). To investigate its potential effects in our study, we tested whether CpG sites where a large proportion of the variance in DNA methylation levels was explained by species identity tended to be sites with methylation levels that strongly differ across blood cell types.

Because no analyses of cell type-specific DNA methylation have been conducted in baboons, we drew on publicly available DNA methylation data generated from pure populations of $CD4^+$ (helper) T cells, $CD8^+$ (cytotoxic) T cells, natural killer cells, B cells, monocytes, and granulocytes from human whole blood (Jaffe and Irizarry 2014). Because these data were generated on Illumina 450K Infinium arrays, relatively few CpG sites overlapped with our data set. However, for the 9,627 CpG sites that were profiled in both the Jaffe data set and in our samples, we considered a CpG site to be differentially methylated by cell type if (i) a significant (10% FDR) difference among cell types was reported by Jaffe & Irizzary (using an ANOVA), and (ii) average DNA methylation levels differed by at least 10% between the two most prevalent white blood cell types (granulocytes and $CD4^+$ T cells, which together account for more than two-thirds of nucleated human blood cells on average and are similarly common in baboons).

Out of the 9,627 sites in our data set that are also on the 450K array, 1,701 met these criteria. We compared the distribution of ANOVA test statistics between sites with and without cell type-specific effects using a Kolmogorov-Smirnov test, and found that sites with cell type-specific effects were not enriched for taxon-specific differences (identified using an ANOVA; two-sided K-S test, $D=0.02$, $p = 0.373$). This analysis suggests that our results are not primarily driven by the effects of cell type heterogeneity. Furthermore, sites with cell type-specific methylation were not enriched for signatures of directional selection using Ornstein-Uhlenbeck models ($\log_2(OR) = 0.023$, $p = 0.657$) or the heuristic approach ($\log_2(OR) = -0.031$, $p = 0.713$).

Populations within a species may also differ in DNA methylation level (Fraser et al. 2012, Heyn et al. 2013, Dubin et al. 2015). In this study, we included samples from two

populations of anubis baboon (n = 2 and 7) and three populations of hamadryas baboon (n = 5, 7, and 2). In a PCA of anubis baboons, individuals from the same population did not cluster together. Similarly, hamadryas baboons did not cluster by source population (ANOVA on PCs 1-4: $p > 0.10$). We also used ANOVA to test for the PVE explained by population, within species. Population assignment did not explain greater variance in our data set than expected by chance (based on permutation of population assignment within species). Thus, differences between populations do not significantly contribute to variation within this dataset.

**References**


Dubin, M. J., P. Zhang, D. Meng, M.-S. Remigereau, E. J. Osborne, F. P. Casale, P. Drewe, A. Kahles, G. Jean, and B. Vilhjálmsson. 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. eLife 4:e05255.

Fraser, H. B., L. L. Lam, S. M. Neumann, and M. S. Kobor. 2012. Population-specificity of human DNA methylation. Genome Biology 13:R8.

Hansen, K. D., B. Langmead, and R. A. Irizarry. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biology 13:1.

Heyn, H., S. Moran, I. Hernando-Herraez, S. Sayols, A. Gomez, J. Sandoval, D. Monk, K. Hata, T. Marques-Bonet, L. Wang, and M. Esteller. 2013. DNA methylation contributes to natural human variation. Genome Research 23:1363-1372.

Houseman, E. A., W. P. Accomando, D. C. Koestler, B. C. Christensen, C. J. Marsit, H. H. Nelson, J. K. Wiencke, and K. T. Kelsey. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86.

Jaffe, A. E., and R. A. Irizarry. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biology 15:R31.

Lea, A. J., J. Altmann, S. C. Alberts, and J. Tung. 2016. Resource base influences genome-wide DNA methylation levels in wild baboons (Papio cynocephalus). Molecular ecology.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, and Q.-M. Ngo. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315-322.

Liu, Y., K. D. Siegmund, P. W. Laird, and B. P. Berman. 2012. Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biology 13:1.

Nei, M., and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76:5269-5273.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289-290.

Reinius, L. E., N. Acevedo, M. Joerink, G. Pershagen, S.-E. Dahlén, D. Greco, C. Söderhäll, A. Scheynius, and J. Kere. 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7:e41361.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100:9440-9445.

Tung, J., and L. B. Barreiro. 2017. The contribution of admixture to primate evolution. Current Opinions in Genetics & Development 47:61.

Zheng, X. 2012. SNPRelate: parrallel computing toolset for genome-wide association studies. R package version 95.

**Table S2**: Enrichment by genomic context for CpG sites with taxonomically structured DNA methylation and positive selection.

| | Gene Introns | Gene Exons | Promoters | Enhancers | CpG Islands | CpG Shores | Unannotated Region |
|---|---|---|---|---|---|---|---|
| **Interspecific differences**[1] | 0.265 ($2.83x10^{-34}$) | -0.569 ($4.00x10^{-52}$) | -0.183 ($3.30x10^{-5}$) | -0.513 ($4.24x10^{-35}$) | -0.422 ($3.08x10^{-38}$) | -0.255 ($2.92x10^{-20}$) | 0.097 ($1.24x10^{-6}$) |
| **Positive Selection**[2] | -0.185 ($1.31x10^{-3}$) | 0.564 ($2.16x10^{-8}$) | 0.506 ($1.35x10^{-5}$) | 0.317 ($3.28x10^{-3}$) | 0.704 ($3.46x10^{-16}$) | 0.491 ($6.92x10^{-11}$) | -0.325 ($1.18x10^{-8}$) |

[1] Enrichment for CpG sites with clade or species level differences in DNA methylation (taxonomically structured sites; n=20,360) relative to the background set of CpG sites that were not constitutively hyper- or hypo-methylated (n = 756,262). Values in each cell give the $\log_2$odds ratio (top) and p-value (bottom) from a Fisher's Exact Test.

[2] Enrichment for CpG sites where variation in DNA methylation is best explained by positive selection (on an individual baboon species using the heuristic method or on a baboon clade using either the heuristic method or OU approach; n = 4,901), relative to the background set of CpG sites with clade or species level differences in DNA methylation. Values in each cell give the $\log_2$odds ratio (top) and p-value (bottom) from a Fisher's Exact Test.

**Table S3**: Enrichment by genomic context for CpG sites with taxonomically structured DNA methylation, stratified by level of taxonomic structure.

| | Gene Introns | Gene Exons | Promoters | Enhancers | CpG Islands | CpG Shores | Unannotated Regions |
|---|---|---|---|---|---|---|---|
| **Genus**[1] | 0.154 ($6.53x10^{-59}$) | -0.311 ($1.08\ x10^{-93}$) | -0.133 ($2.97x10^{-12}$) | -0.119 ($3.01x10^{-13}$) | -0.125 ($2.74\ x10^{-21}$) | -0.144 ($3.20x10^{-35}$) | 0.026 ($1.99x10^{-3}$) |
| **Clade** | 0.229 ($3.89x10^{-17}$) | -0.365 ($5.46x10^{-16}$) | -0.013 (0.417) | -0.463 ($2.63x10^{-19}$) | -0.215 ($3.62x10^{-8}$) | -0.107 ($8.92x10^{-4}$) | 0.003 (0.451) |
| **Species** | 0.344 ($2.17x10^{-29}$) | -0.909 ($8.67x10^{-58}$) | -0.449 ($3.28x10^{-11}$) | -0.680 ($5.56x10^{-29}$) | -0.803 ($8.17x10^{-59}$) | -0.506 ($3.96x10^{-35}$) | 0.344 ($2.17x10^{-29}$) |

[1] Enrichment for CpG sites with differences in DNA methylation between macaques and baboons (row 1), northern and southern clade baboons (row 2), and baboon species within a clade (row 3) (ANOVA 10% FDR). Values in each cell give the $\log_2$odds ratio (top) and p-value (bottom) from a Fisher's Exact Test.

**Table S4**: Enrichment by genomic context for CpG sites with evidence for positive selection, stratified by modeling approach and level of taxonomic structure.

| | Gene Introns | Gene Exons | Promoters | Enhancers | CpG Islands | CpG Shores | Unannotated Regions |
|---|---|---|---|---|---|---|---|
| **Species: Heuristic** | -0.166 ($5.20x10^{-3}$) | 0.355 ($1.34x10^{-3}$) | 0.460 ($2.97\ x10^{-4}$) | 0.242 ($4.88x10^{-2}$) | 0.616 ($2.78x10^{-11}$) | 0.435 ($4.90x10^{-8}$) | -0.280 ($5.34\ x10^{-6}$) |
| **Clade: Heuristic** | -0.277 ($1.08x10^{-2}$) | 0.938 ($1.25x10^{-8}$) | 0.710 ($3.75x10^{-4}$) | 0.328 (0.103) | 1.106 ($2.27x10^{-15}$) | 0.619 ($9.71x10^{-7}$) | -0.525 ($3.55x10^{-7}$) |
| **Clade: OU models** | -0.061 (0.170) | 0.245 ($3.19x10^{-3}$) | 0.022 (0.843) | 0.244 ($7.69x10^{-3}$) | 0.180 ($1.08x10^{-2}$) | 0.203 ($5.24x10^{-4}$) | -0.064 (0.130) |

[1] Enrichment for CpG sites with evidence for positive selection, stratified by method and taxonomic level, relative to the set of sites with taxonomic structure in DNA methylation (n = 20,360). Values in each cell give the $\log_2$odds ratio (top) and p-value (bottom) from a Fisher's Exact Test.

**Table S5:** Proportion of simulated sites assigned to alternative OU models when positive selection affects multi-species lineages.

| | | Assigned to | | |
|---|---|---|---|---|
| | | Genetic Drift | Stabilizing Selection | Directional Selection |
| **Simulated under** | Brownian Motion | 79.16% | 1.65% | 19.19% |
| | Stabilizing Selection | 78.49% | 4.82% | 16.69% |
| | Shift: Northern Clade | 36.36% | 2.62% | 61.02% |
| | Shift: Southern Clade | 41.47% | 2.55% | 55.98% |
| | Shift: Anubis-Guinea | 38.88% | 2.81% | 58.31% |

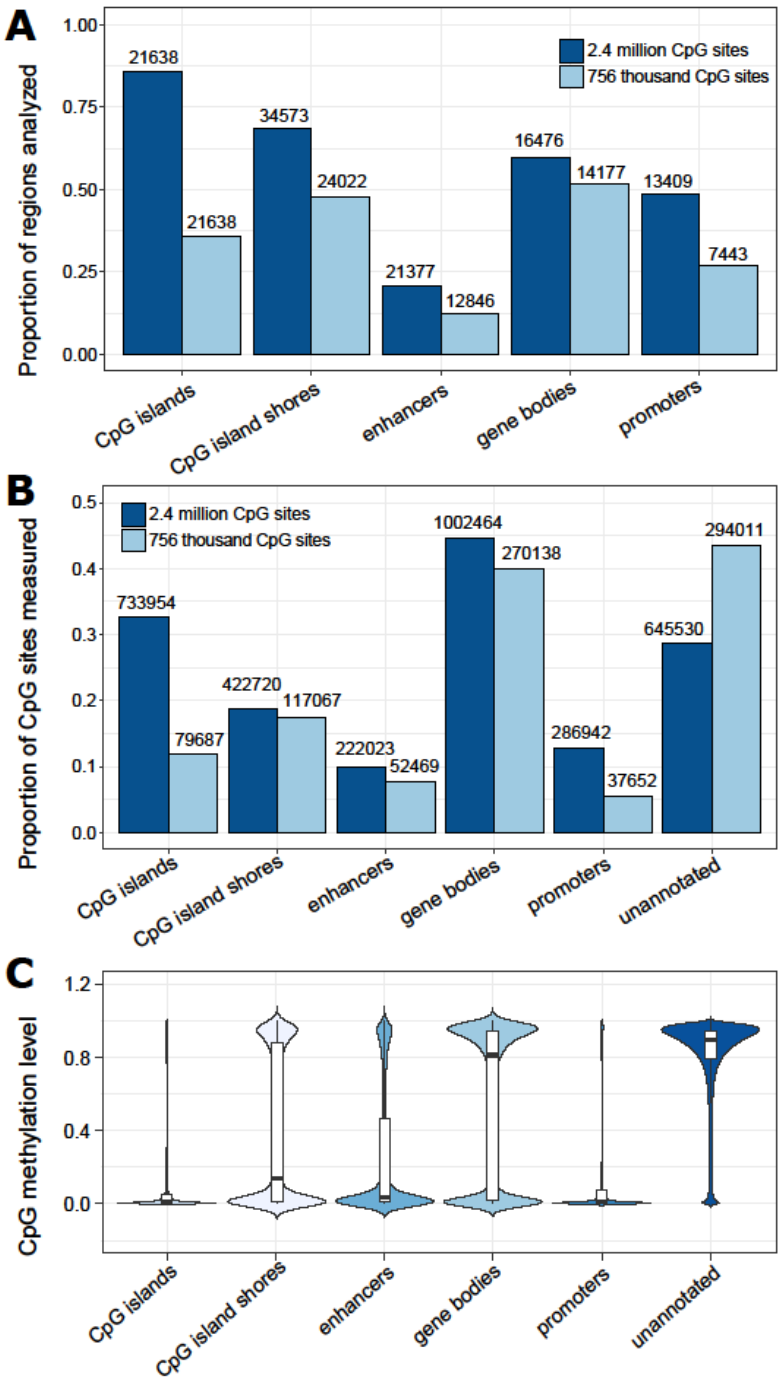


**Figure S1: Properties of the RRBS data set.** Dark blue bars show all sites with 5x coverage (n = 2.4 million sites); light blue bars show sites filtered for mean methylation between 10% and 90% (n = 756k sites). **(A)** Proportion of annotated features in the baboon genome for which at least one CpG site was analyzed. Numbers above each bar represent the number of features tagged in our data set. **(B)** Proportion of total CpG sites analyzed that fell in each genomic region. Numbers above each bar represent the number of CpG sites included from each category. **(C)** Violin plots showing the distribution of mean DNA methylation levels for CpG sites located in each genomic context. White box plots show the interquartile range and median (black bar).

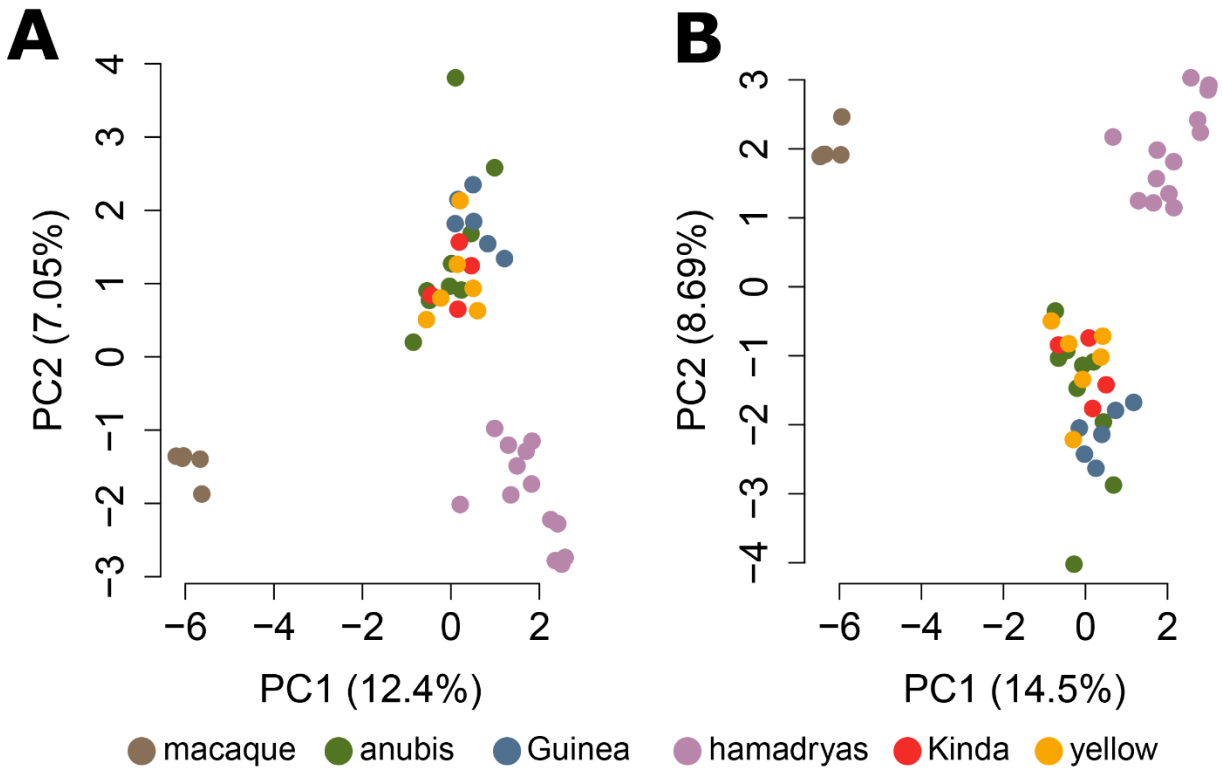


**Figure S2: Principal component analyses of DNA methylation levels, including macaques.** PCA was performed on all measured CpG sites (n = 2.45 million sites; panel **A**) and for sites filtered for mean methylation between 10% and 90% (n = 756k sites; panel **B**). In both cases, macaques separate from all baboon samples along the first principal component of variation; subsequent PCs separate baboon species.

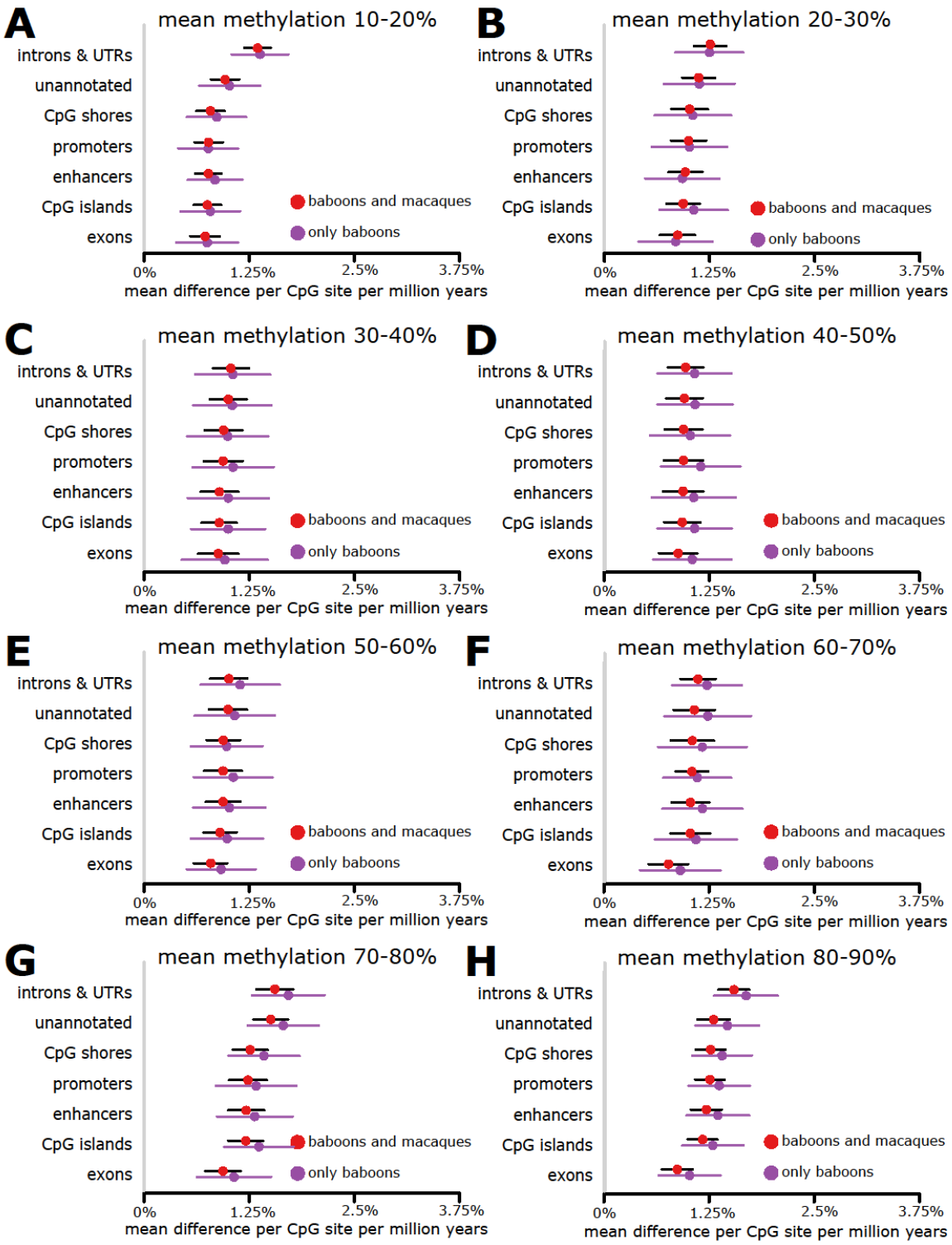


**Figure S3: Mean rate of change in DNA methylation levels, controlling for mean methylation.** Estimated mean rate of change in DNA methylation levels per million years, stratified by genomic context. Data are split by mean methylation level (panels **A-H**).

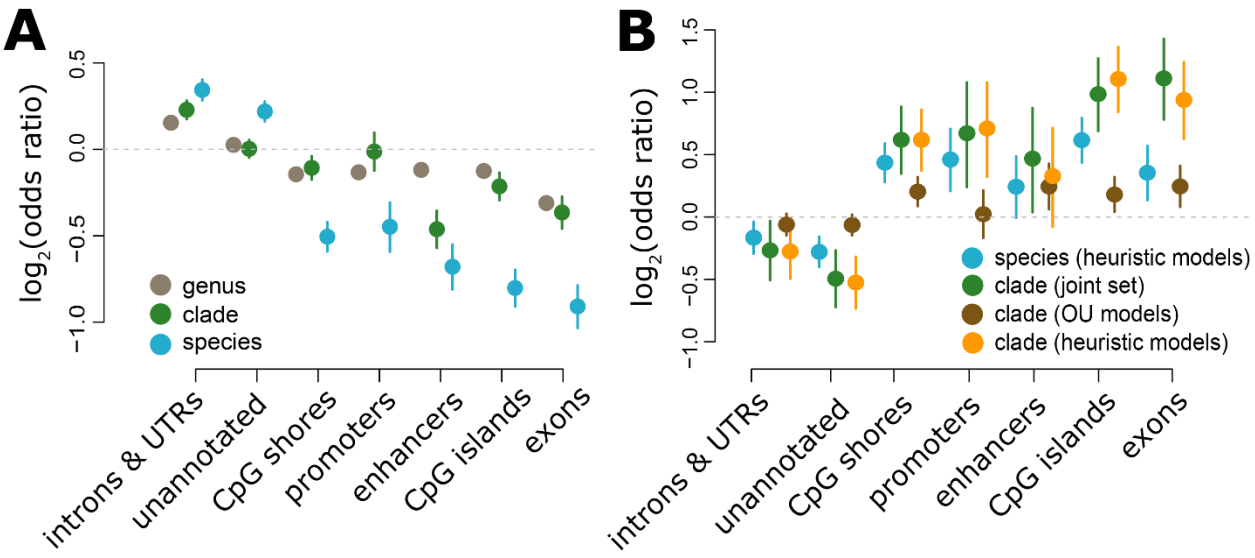


**Figure S4: Enrichment by genomic context for sites with taxonomically structured DNA methylation levels. (A)** Enrichment by genomic context for sites with taxonomically structured DNA methylation levels at the level of genus, baboon clade, and baboon species. **(B)** Enrichment by genomic context for candidate positively selected CpG sites relative to the set of taxonomically structured sites. The three values presented for clade-level selection refer to results from Ornstein-Uhlenbeck models, the heuristic approach, and the intersection of both approaches.

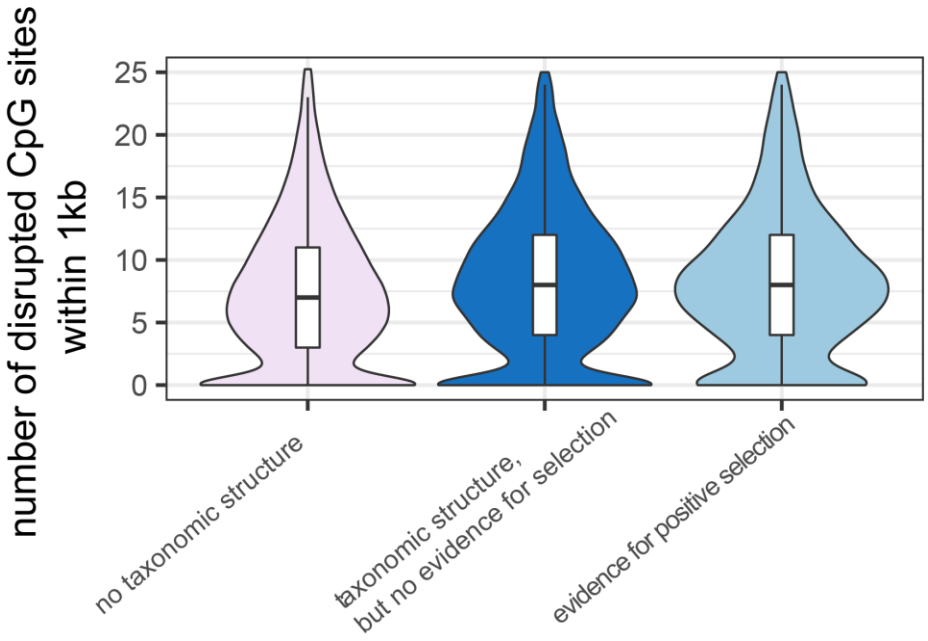


**Figure S5: CpG sites with taxonomically structured variation fall near a larger number of disrupted CpG sites.** The number of nearby (within 1kb) CpG sites for which one version of the sites carries an intact CpG site, but the alternate version leads to a loss of the CpG site, for each CpG sites. CpG sites with taxonomic structure (blue) were nearby more disrupted CpG sites than CpG sites without taxonomic structure (pink) (logistic regression: $\beta=0.0246$, $p=3.33\times10^{-133}$). Within taxonomically structured sites, there was no difference based on evidence for positive selection (dark vs light blue) ($\beta=1.93\times10^{-3}$, $p=0.471$). The y-axis is truncated at 25, although each category includes a small number of sites with up to ~80 nearby disrupted CpG sites.

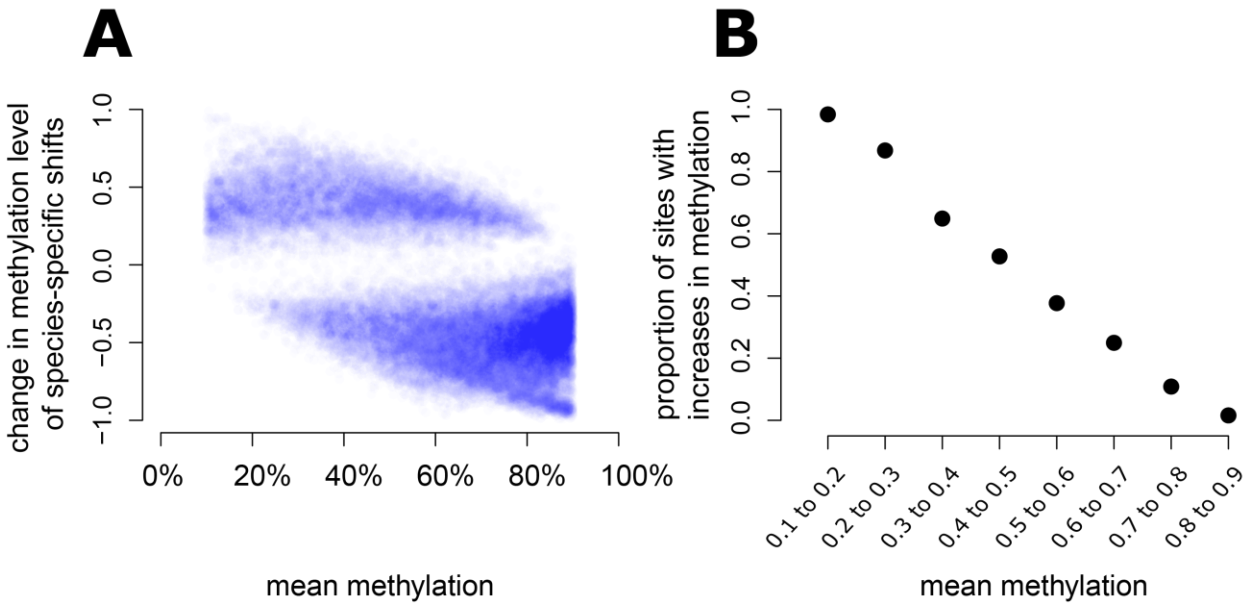


**Figure S6: Directionality of shifts in DNA methylation is accounted for by mean methylation level. (A)** Difference in mean methylation level between species that experienced a shift in DNA methylation level and mean DNA methylation level across all baboons (y-axis), plotted against mean methylation across all baboons (x-axis). Only CpG sites associated with species-specific shifts are shown (no sites with mean methylation levels <10% or >90% were included in this analysis). Larger shifts tend to occur at sites with more extreme mean methylation levels. **(B)** Proportion of species-specific shifts in DNA methylation that involved hypermethylation on the shifted lineage, stratified by mean methylation level across all baboons. Sites that are hypomethylated exhibited increases in methylation on shifted lineages, whereas sites that are hypermethylated exhibited decreases in methylation on shifted lineages.

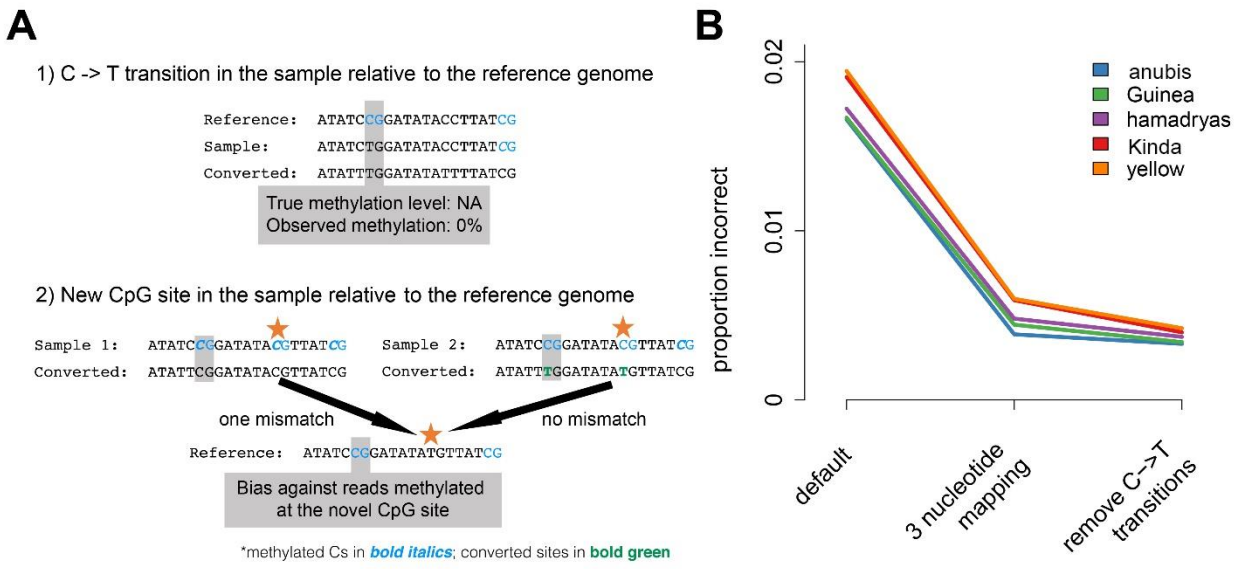


**Figure S7. Heterospecific mapping biases. (A)** There are two cases in which mapping to a heterospecific individual produces biases in DNA methylation level estimates. In the first case, a C →T transition at a CpG site results in the site being observed as hypomethylated even though no CpG site exists. In the second case, a T → C transition creates a novel CpG site. Because unmethylated cytosines are converted to match the reference (Sample 2), there is only a mismatch to the reference genome if the new CpG site is methylated. This results in a mapping bias towards reads where the new CpG site is unmethylated, which can affect estimates of DNA methylation at nearby (reference) CpG sites if methylation levels at the reference CpG sites and the new CpG site are correlated. **(B)** The proportion of CpG sites for which simulated DNA methylation levels were influenced by a mapping bias. 3 nucleotide mapping and removing C → T transitions reduces the proportion of incorrect estimates, which tend to be higher for southern clade species than northern clade species (the reference genome was sequenced from a northern clade baboon).

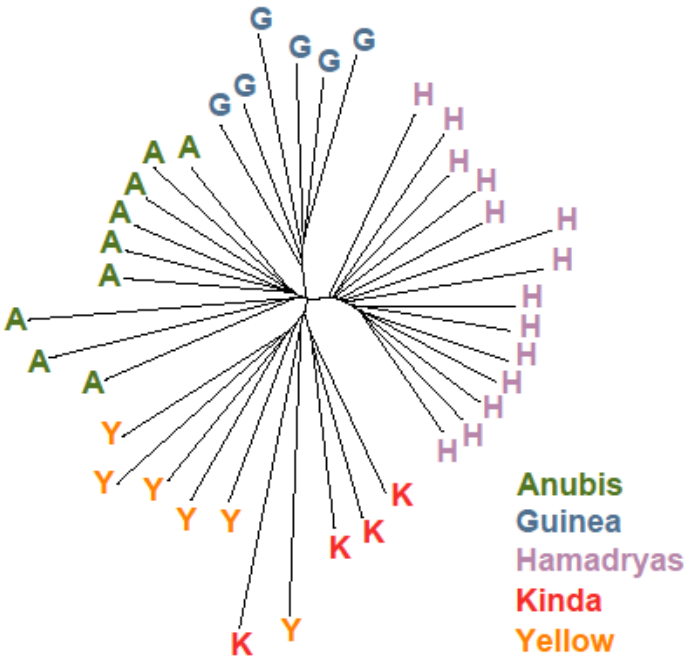


**Figure S8. Structure of genotype data from RRBS.** Unrooted neighbor joining tree performed on 49,607 baboon SNPs that were called from RRBS data. With the exception of a kinda and a yellow individual that were missing a large proportion of genotype calls (>55%), individuals of each species cluster together and the relationship between species recapitulates the known baboon phylogeny (Rogers et al. *in press*).