Supporting Information

Variation in the molecular clock of primates

Priya Moorjani, Carlos Eduardo G. Amorim, Peter F. Arndt, Molly Przeworski

Table of contents

Supplementary notes

Note S1: Methods and Materials.
Note S2: High coverage human genome.
Note S3: Analysis of Enredo-Pecan-Ortheus (EPO) dataset.

Supplementary Figures

Figure S1: Sperm methylation profiles at CpG sites.
Figure S2: Distribution of CpG and non-CpG G/C sites across the human genome.
Figure S3: Comparison of substitution rates in hominoids and Old World Monkeys (OWMs) using alternate topologies.
Figure S4: Comparison of substitution rates in hominoids and New World Monkeys (NWMs) using alternate topologies.
Figure S5: Phylogenetic tree for the six primates in EPO dataset.
Figure S6: Comparison of substitution rates in hominoids and OWMs using different datasets.
Figure S7: Variance among lineages for distinct substitution types, estimated from different datasets.
Figure S8: Effect of biased gene conversion across lineages estimated for different datasets.
Figure S9: Comparison of substitution rates in human and chimpanzee using Phylofit.
Figure S10: Comparison of substitution rates in human and gorilla using Phylofit.
Figure S11: Comparison of substitution rates in human and chimpanzee using the maximum likelihood approach.
Figure S12: Comparison of substitution rates in human and gorilla using the maximum likelihood approach.
Figure S13: Mutation spectrum across primates.

Supplementary Tables

Table S1: Online source of annotation for transposable elements, coding exons, CpG Islands (CGI), and conserved sites.
Table S2: Life history traits in primates.
Table S3: Autosomal substitution rates on the human lineage for different time depths and using different filters.
Table S4: Correlation in life history traits across primates.

References

Note S1: Materials and Methods

Data sets and filtering. We used the following datasets for our analysis:
(a) **Multiz:** A 12-primate whole genome sequence alignment, with mouse as an outgroup, which is part of a 100-way mammalian phylogeny, mapped using Multiz (1).

(b) **Enredo-Pecan-Ortheus (EPO):** A seven primate whole genome alignment, mapped using the EPO pipeline (2). We removed duplications using the mafDuplicateFilter from mafTools package (3). This software identifies any duplicated region in the alignment block and only retains the sequence with the highest similarity to the consensus sequence.

(c) **High coverage hominoid dataset**: We generated pairwise sequence alignments of high coverage genomes for human, chimpanzee and gorilla, consisting of a human (of European ancestry) that we sequenced in collaboration with Carole Ober (Department of Human Genetics, University of Chicago) (Note S2), a chimpanzee (Ind-D from (4)) and a gorilla (Delphi from (5); data kindly provided by Tomas Marques-Bonet, Institut Biologia Evolutiva, Universitat Pompeu Fabra / Spanish National Research Council (CSIC)). These genomes were mapped to the orangutan reference genome (ponAbe2) (6), which should be equidistant to humans and extant African great apes (assuming no variation in substitution rates), using bwa-mem (7) with default parameters and the multi-threading option (-t). The coverage after mapping was as follows: human = 30.21, chimpanzee = 31.23 and gorilla = 32.75. Because library information was not available for all primates, to ensure symmetry in our treatment of all primate genomes, we did not remove optical duplicates. Single nucleotide polymorphisms (SNP) in each high-coverage diploid genome were called using samtools mpileup (version: 0.1.18-dev) (7) with the -B option (to reduce the number of false SNPs called due to misalignments). The bam files were converted to fasta format using BCFtools and seqtk (part of samtools) and only sites that had a minimum quality score of 30 were retained for further analysis (-q30). As we need haploid genomes in our inference procedure, for each polymorphic site in the high coverage genomes, we randomly sampled one allele, thereby generating a pseudo-haploid genome for each species. These high coverage and high quality fasta files were used for pairwise comparisons of human-chimpanzee and human-gorilla genomes, with the orangutan reference genome used as the outgroup.

For the three datasets, we filtered out missing data, i.e., any base pair that was aligned to a gap or a missing site in at least one of the primate species. To consider putatively neutral sites, we limited our analyses to the non-coding, non-conserved and non-repetitive regions of the genome (see Table S1 for the source of all annotations used). For each primate species, we excluded sites with the following annotations:
(a) Conserved elements annotated using phastCons (8) based on the multiple alignments of 46 primates (9). These annotations were downloaded from UCSC browser (track: phastConsElements46wayPrimates).
(b) Coding exons based on the NCBI RNA reference sequences collection annotation or equivalent. These annotations were downloaded from UCSC browser (track: RefSeq Genes).

(c) Transposable elements. As the levels of methylation are higher for repetitive regions than non-repetitive regions of the genome (10), which could lead to differences in mutation rates, we removed the repetitive regions including interspersed nuclear elements (LINE and SINE), DNA repeat elements and Long Terminal Repeat elements identified using RepeatMasker (11).

In some cases, we also excluded sites within CpG islands (CGI). Transitions at CpG sites are thought to primarily occur due to spontaneous deamination at methylated cytosines. However, within CGI, most CpGs are hypomethylated (12). As an illustration, comparison of sperm methylation profiles in humans from (13) showed that only 7.5% of CpG sites in annotated CGI have a methylation level of greater than or equal to 40% whereas the vast majority (84.6%) of CpG sites outside CGI have similar or greater methylation levels (Figure S1). To focus on a more homogeneous set of methylated CpGs, we therefore excluded CGI from the analysis, unless otherwise specified. CGI annotations were downloaded from UCSC browser (track: CpG Islands) (14).

Estimating substitution rates. We used Phylofit (15) to estimate autosomal substitutions for the three datasets described above. To access the robustness of the estimates from Phylofit, we also used an alternative maximum likelihood based approach from (16) for the high coverage hominoid genomes. Both methods require as input the topology of the phylogenetic tree for the species represented in the analysis, which were subsets of the primates included in the Multiz, EPO or the high coverage hominoid dataset. Because these methods assume a single tree for all sites (i.e., ignore the possibility of incomplete lineage sorting), for species pairs with known and non-negligible incomplete lineage sorting, such as human/chimpanzee/gorilla and human/gibbon/orangutan (17), we considered only one of the two lineages in a given analysis.

Phylofit (15) analysis was performed with the expectation maximization algorithm (option -E) with medium precision for convergence. For both internal and external branches, Phylofit outputs both the overall branch lengths (based on all substitutions), accounting for recurrent substitutions at a site, and "posterior counts", i.e., posterior mean of substitutions of each type on each branch, summed across all sites (option -Z). We used the U2S substitution model (the general unrestricted dinucleotide model with strand symmetry) with overlapping tuples to estimate lineage-specific CpG substitution rates and UNREST (the general unrestricted single nucleotide model) to estimate the non-CpG substitution rates. To ensure that the branch lengths across U2S and UNREST are comparable, we ran UNREST with fixed branch lengths that were estimated using U2S.

In running Phylofit multiple times, we observed that a subset of the runs, often with substantially lower likelihoods, returned different point estimates for the overall branch lengths. We interpret this finding as reflecting the fact that the method sometimes returns values for a local peak in the likelihood surface. To circumvent this problem, we ran Phylofit ten times with different seeds (using -r and -D options) and report the estimate for the run with the highest likelihood. We note, however, that even estimates from runs with lower likelihoods were fairly similar and the posterior counts were essentially identical.

We used the posterior counts from Phylofit to estimate the number of substitutions involving transitions and transversions for the following types of sites: ancestrally A or T sites (referred to as A/T), ancestrally G or C sites (G/C), ancestrally CG dinucleotides (CpG) and ancestrally G or C sites that are not part of a CG dinucleotide (non-CpG G/C). Specifically, for each mutation type, we estimated the divergence from an internal node to the terminal node as the mean posterior number of positions at which the ancestral allele A_1 (at the internal node) is inferred to have been substituted to allele A_2 (at the terminal node) on that lineage divided by the total count of ancestral alleles A_1 at that internal node. In doing so, we are implicitly assuming a single mutation from A_1 to A_2, thereby making a parsimony assumption. To study the effects of biased gene conversion, we similarly estimated the substitution rates for strong (S; G/C) and weak (W; A/T) mutations in different substitution contexts (CpG or non-CpG).

For the high coverage hominoid analysis (dataset (c)), we ran Phylofit five times with five different seeds (using -r and -D options) and report the estimate for the run with the highest likelihood. Additionally, we also used the maximum likelihood based approach from (16). This approach uses a probabilistic model for sequence evolution and assumes that all nucleotide substitutions except those occurring in a CpG context evolve independently. Thus there are 6 parameters in a reverse complement symmetric analysis and 12 parameters if the complement strands evolve with different rates. Substitutions at C and G in the CpG context have their own rates, which yields three or six additional parameters in the reverse complement symmetric setting or non-reverse complement symmetric setting, respectively. To account for context dependence of the adjacent nucleotides, the maximum likelihood approach computes the evolution of tri-nucleotides. Unlike Phylofit, the maximum likelihood approach does not assume that the nucleotide substitution process is in stationary state. This method was run with multi-threading and strand-asymmetry option to estimate the rate of 12 context-free substitutions (A->[C/T/G], T->[A/C/G], non-CpG C->[A/T/G] and non-CpG G->[A/T/G]) and six CpG substitutions (two CpG transitions: CG->[CA/TG] or four CpG transversions: CG->[CC/GG/CT/AG]). To obtain estimates of the number of transitions and transversions for different ancestral contexts (A/T, CpG and non-CpG G/C), we estimated a weighted average of the rates across symmetric classes of substitutions using the counts of the nucleotides in the orangutan genome for normalization.

Estimating the root-leaf variance. For each substitution type, we constructed a phylogenetic tree using the lineage-specific substitution rates estimated by Phylofit for the Multiz and EPO datasets. We computed the root-leaf distance using the *R* package *adephylo* (18). Following (19), we considered the variance in the root to leaf distance after normalizing by the mean distance. We note that while this procedure results in counting some ancestral branches more than once, the analysis performed with single representatives from each species group yields qualitatively similar results (not shown).

Assessing the significance of branch length differences in pairwise comparisons. To test if the branch lengths estimated by Phylofit differ between two species, we used a likelihood ratio test where the null model is that the number of substitutions on the

branch leading to both species are equal and the alternative that they were not equal. Thus, the likelihood ratio statistic

$$\Delta = 2[n_1 \log\left(\frac{n_1}{0.5n}\right) + (n - n_1)\log\left(\frac{n - n_1}{0.5n}\right)]$$

should be approximately $\chi^2(df = 1)$, where n_1 is the number of substitutions leading to species$_1$ and n_2 to species$_2$ and $n = n_1+n_2$.

Phylogenetically independent contrast analysis: We tested the correlation between generation time and non-CpG substitution rates using the phylogenetically independent contrasts (*pic*) method described by Felsenstein (20) that is implemented in the *R* package *ape* (21). Because of the quasi-clocklike behavior of CpG transition rates, we use these substitutions to specify branch lengths for the phylogeny. Generation time estimates assumed for all extant species are shown in Table S2.

Modeling yearly mutation rates. To estimate the average yearly mutation rates (μ_y) for a given set of life-history traits, we used the mutational model from (22). In this model, the mutation rate per year is given by:

$$\mu_y = \frac{\mu_F + C_M - I(D_M/\tau)\,(G - P - I)}{G_F + G_M}$$

where μ_F is the female mutation rate per generation, C_M is the expected number of mutations that occurred pre-puberty, I is the gestation time, $\tau = \frac{365}{SECL}$ is the number of spermatogonial stem cell divisions each year for a given rate of spermatogenesis (measured by estimating the seminiferous epithelium cycle length (SECL)), D_M is the expected number of mutations per spermatogenic division, and D_M/τ is the expected mutation rate per year in males. P is the onset of puberty in males and G_F, G_M, G refer to the mean age of reproduction in females, males and the average across both species, respectively.

Following (22), and despite considerable uncertainty in these estimates (23), we assumed mutational parameters to be $C_M = 6.13 \times 10^{-9}$, $D_M = 3.33 \times 10^{-11}$ and $\mu_F = 5.42 \times 10^{-9}$per bp (24). Parameter values for life-history traits used for all species are shown in Table S2.

Estimating average divergence and split times in hominines using CpG transitions. We estimated the divergence time between human-chimpanzee and human-gorilla using substitutions involving transitions at CpG sites (outside CGI), as:

$$t_{divergence} = \frac{X_{CG\rightarrow TG/CA}}{\mu_{CG\rightarrow TG/CA}}$$

where $X_{CG\rightarrow TG/CA}$ is the number of transitions that occurred at CpG sites on the human lineage since the split from the common ancestor (i.e. either the human-chimpanzee or human-gorilla common ancestor) and $\mu_{CG\rightarrow TG/CA}$ is the per year mutation rate for CpG

transitions. We estimated $X_{CG\rightarrow TG/CA}$ from the mean posterior counts reported by Phylofit; in turn, the estimate of $\mu_{CG\rightarrow TG/CA}$ (=3.9x10^{-9} per base pair per year) was obtained by dividing the per generation mutation rate at CpG transitions (= 1.12x10^{-7} per base pair per generation) in (24) by the mean parental age in that study (28.4 years), which is appropriate if the number of CpG transitions increase strictly proportionally to age (as they must if clock-like (25)) .

Assuming an instantaneous split between human and chimpanzee, $t_{divergence} = t_{split} + t_{MRCA}$. Further assuming a panmictic, constant size population, $t_{MRCA} = 2N_aG$, where N_a is effective population size of the ancestral population and G the generation time in the ancestral population of humans and apes. Therefore:

$$t_{split} = \frac{X_{CG\rightarrow TG/CA}}{\mu_{CG\rightarrow TG/CA}} - 2N_aG$$

Previous studies suggest that $N_a = 5N_h$ (5, 26) where N_h is the effective population size in contemporary humans. We estimated N_h as $\pi_{CG\rightarrow TG/CA}/\ 4\mu_{CG\rightarrow TG/CA}$, where $\pi_{CG\rightarrow TG/CA}$ is the average diversity level observed at transitions at CpG sites across 13 diverse human populations (27, 28).

Web resources. Datasets used for the analysis can be downloaded form:
http://przeworski.c2b2.columbia.edu/index.php/softwaredata/

Note S2: High coverage human genome. We sequenced one individual of European ancestry, in collaboration with Carole Ober (Department of Human Genetics, University of Chicago). This individual provided informed consent for participation in the study. The project was approved by Institutional Review Boards at the University of Chicago and Columbia University.

Genomic DNA was extracted from blood and libraries were generated with the Illumina PCR-free library making kit. Briefly, 1µg of DNA was extracted and sheared into fragments using sonication. The resulting fragments were end repaired, a single adenosine overhang was added and indexed paired-end adaptors were ligated. Gel electrophoresis was performed to select libraries with insert sizes of approximately 350 bp in size, which was validated using quantitative PCR. The resulting libraries were sequenced using Illumina HiSeq2500 (v3 chemistry) to generate paired-end reads. We generated ~89 Gb of sequencing data (~30x coverage). Mapping and alignment were done using samtools as described in Note S1.

Sequence data are available through dbGaP:
https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000185.v3.p1

Note S3: Analysis of Enredo-Pecan-Ortheus (EPO) dataset. To test the robustness of our inferences, we repeated the analysis with the EPO dataset containing seven primates (human, chimpanzee, gorilla, orangutan, rhesus macaque, baboon and marmoset). Due to concerns of incomplete lineage sorting between chimpanzee/gorilla/human (17), we used human and chimpanzee and excluded gorilla from further analysis. After filtering putatively non-neutral sites and removing missing data, we analyzed approximately 745 Mb of whole genome sequence alignment. To allow for direct comparison with the Multiz dataset, we repeated our main analysis with the same smaller subset of species available for the EPO dataset. Due to challenges in accurately reconstructing the ancestral state for outgroup species, here marmoset, substitution rates in NWM could be underestimated and hence we do not include comparisons of hominoids and NWM for this dataset.

We applied Phylofit to estimate the substitution rates across all species (Figure S5) and found that substitution rates on lineages leading from the hominoid-OWM ancestor to hominoids are on average 2.81% (range: 2.75- 2.88% across species), whereas rates on lineages leading to OWM are on average 3.57% (range: 3.565- 3.570%), 1.27-fold higher. These estimates are lower than results reported in the main text, likely as we are using a smaller subset of species. Indeed, we obtained similar estimates when analyzing a similar subset of species in the Multiz dataset, obtaining substitution rates that are 1.28-fold faster in OWM compared to hominoids. We also repeated the main analyses shown in Figure 2 and 3 with the smaller subset of species in the EPO and Multiz dataset (see Figure S6-S8).

Figure S1: Sperm methylation profiles at CpG sites. The distribution of methylation levels at CpG sites inside and outside of annotated CGI. The methylation profiles in human sperm were taken from (13). R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS1.R

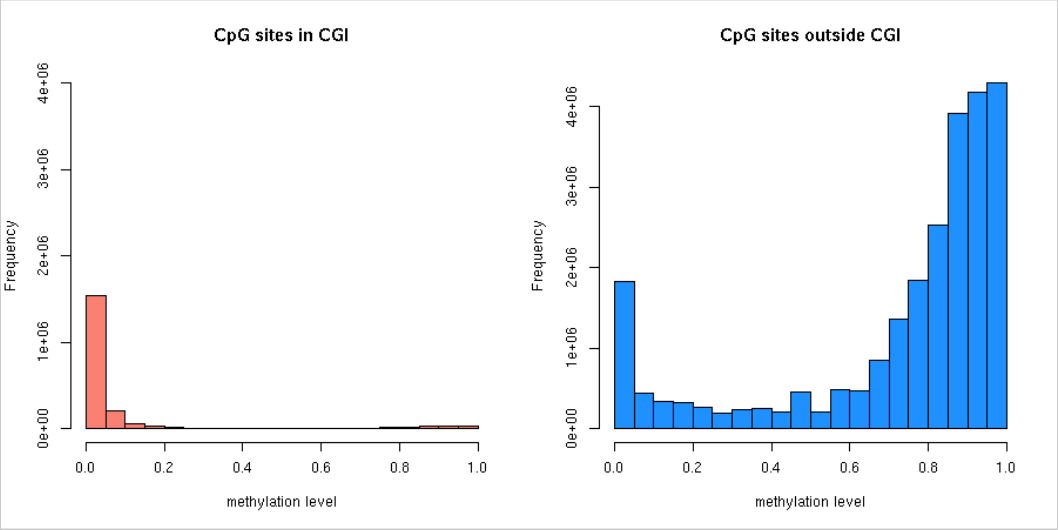


Figure S2: Distribution of CpG and non-CpG G/C sites across the human genome. The proportion of CpG and non-CpG G/C sites in the human genome, as a function of the recombination rate is shown. After filtering non-neutral sites and CGI (see Note S1) in the Multiz dataset, the proportions of CpG and non-CpG G/C sites are 1.60% and 37.9%, respectively. Crossover rates were obtained from the UCSC genome browser track "deCODE Recombination maps: Sex avg" (29), which were estimated in cM/Mb for 10 kb bins and standardized to have an average rate of 1 across the genome. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS2.R

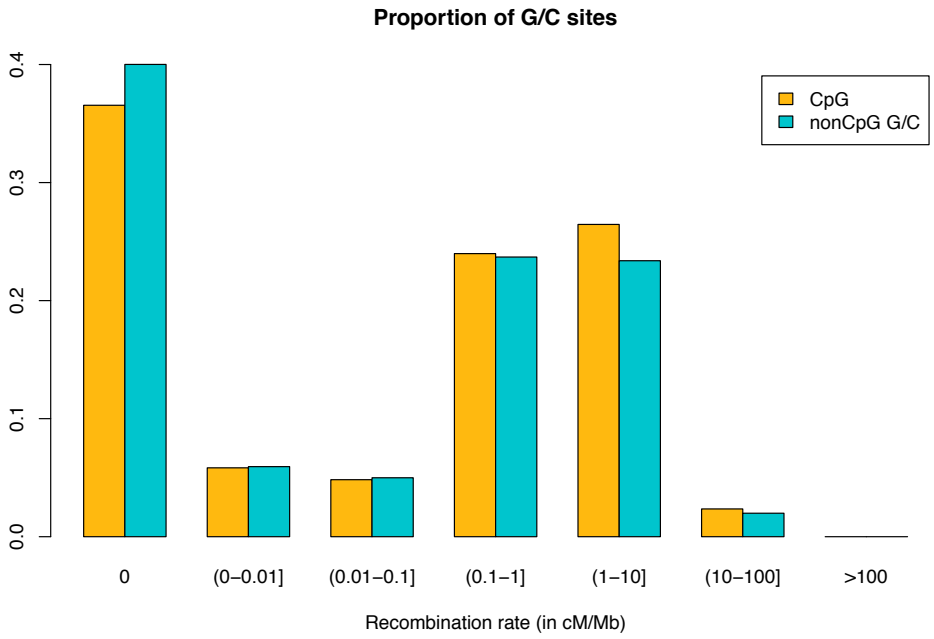


Figure S3: Comparison of substitution rates in hominoids and Old World Monkeys (OWMs) using alternate topologies. Due to concerns about the possible effects of incomplete lineage sorting, we analyzed gorilla and chimpanzee (and gibbon and orangutan) separately. Each sub-figure shows a different set of species and substitution type (transitions at CpG or non-CpG G/C sites). For each topology, we estimated the total branch length from the hominoid-OWM ancestor to each leaf. The branch length from the root to the human tip was set to 1 (the actual value is shown in parenthesis), and other lineages were normalized to the human branch length. Branches from root to hominoids are shown in purple and from root to OWMs are shown in green. The ratio of the average substitution rate from the root to OWMs to the average rate from the root to hominoids is shown as the title for each sub-figure. R code to replicate this figure is available at:
https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS3.R

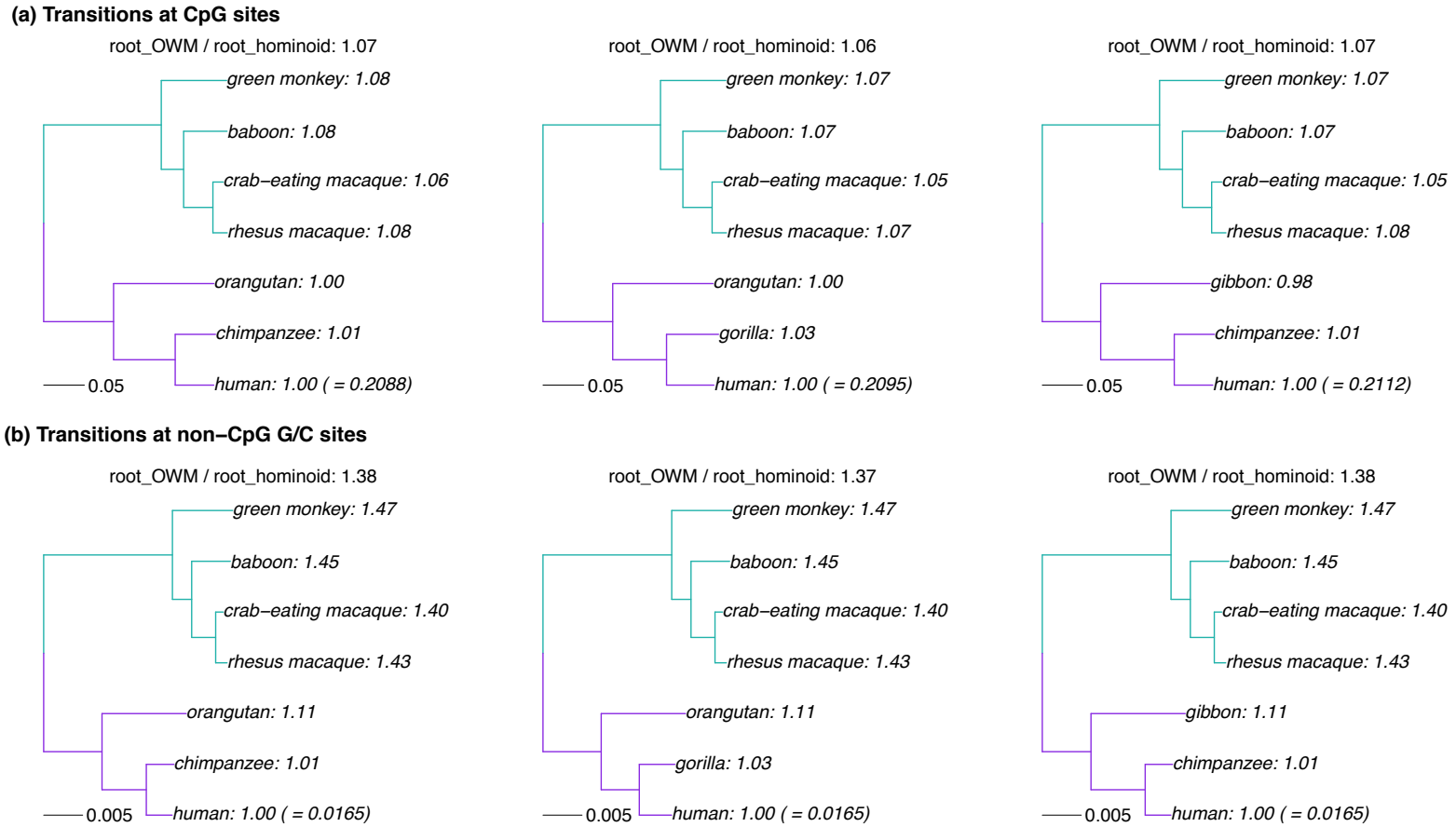


Figure S4: Comparison of substitution rates in hominoids and New World Monkeys (NWMs) using alternate topologies. Due to concerns about the possible effects of incomplete lineage sorting, we analyzed gorilla and chimpanzee (and gibbon and orangutan) separately. Each sub-figure shows a different set of species and substitution type (transitions at CpG or non-CpG G/C sites). For each topology, we estimated the total branch length from the hominoid-NWM ancestor to each leaf. The branch length from the root to the human tip was set to 1 (the actual value is shown in parenthesis), and other lineages were normalized to the human branch length. Branches from root to hominoids are shown in purple and from root to NWMs are shown in green. The ratio of the average substitution rate from the root to NWMs to the average rate from the root to hominoids is shown as the title for each sub-figure. R code to replicate this figure is available at:
https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS4.R

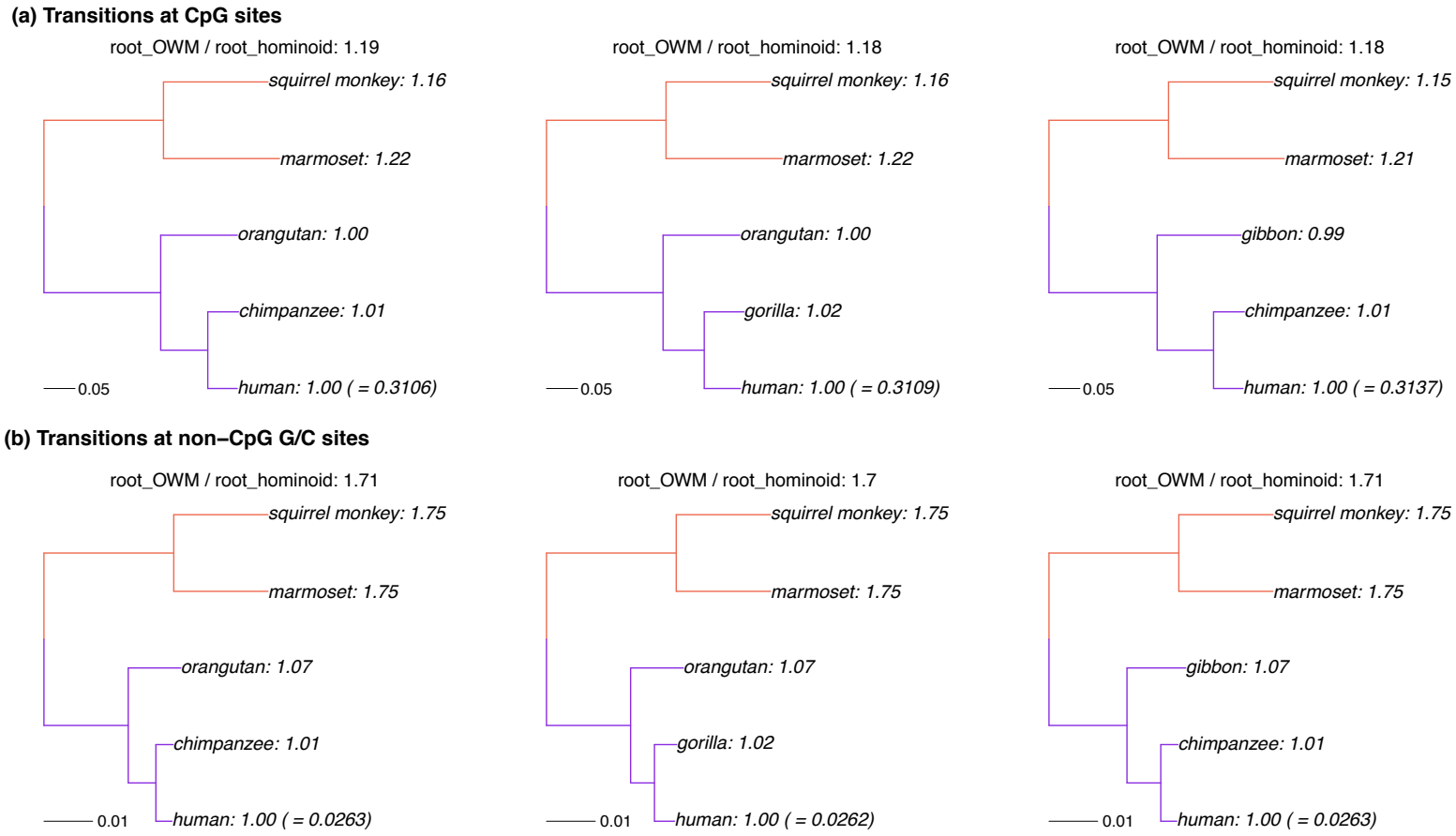


Figure S5: Phylogenetic tree for the six primates in EPO dataset. We estimated neutral substitution rates for six primates from the EPO dataset using Phylofit (see Note S1 for details). Branch lengths reflect the expected number of neutral substitutions per site along each lineage. We excluded gorilla due to concerns about possible effects of incomplete lineage sorting on estimated substitution rates. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS5.R

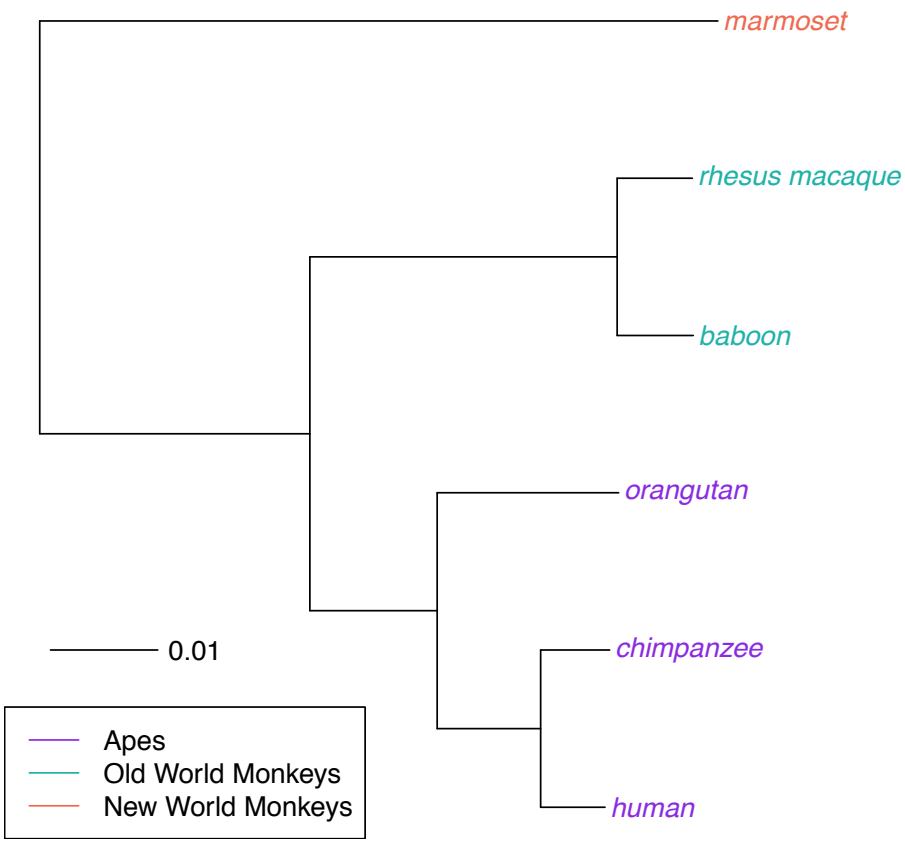


Figure S6: Comparison of substitution rates in hominoids and OWMs using different datasets. For each dataset (Multiz or EPO), we estimated the total branch length from the hominoid-OWM ancestor (root) to each leaf. The branch length from the root to the human tip was set to 1 (the actual value is shown in parenthesis), and other lineages were normalized to the human branch length. Branches from root to hominoids are shown in purple and from root to OWM are shown in green. The ratio of the average substitution rate from the root to OWMs to the average rate from the root to hominoids is shown as the title for each sub-figure, along with the substitution context. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS6.R

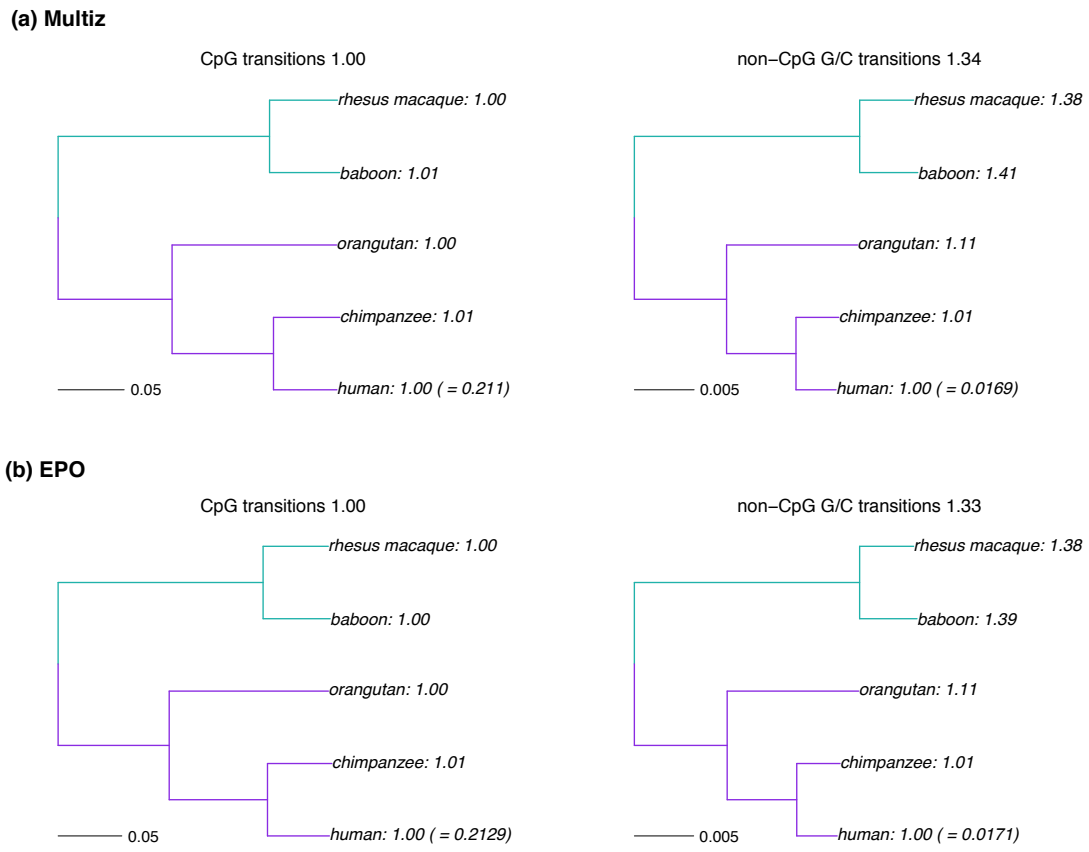
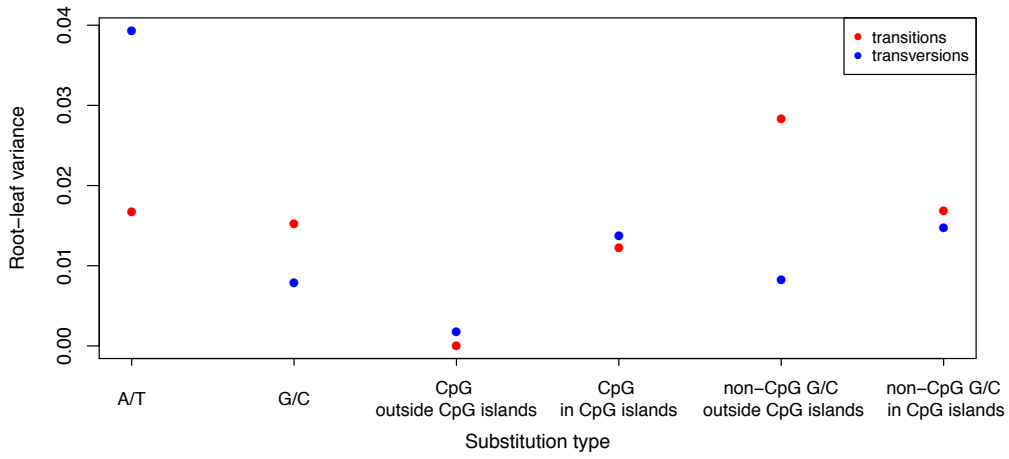


Figure S7: Variance among lineages for distinct substitution types, estimated from different datasets. For each ancestral state and each context shown on the x-axis, we estimated the total branch length from the root to each terminal leaf as the inferred number of substitutions per site, in (a) Multiz and (b) EPO dataset. We then computed the variance in the normalized root to leaf distance across five primates (human, chimpanzee, orangutan, rhesus macaque and baboon). This figure differs from Figure 2A, as it uses fewer species in the Multiz dataset to match the set of species (hominoids and OWMs) available in the EPO dataset. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS7.R

(a) Multiz



(b) EPO

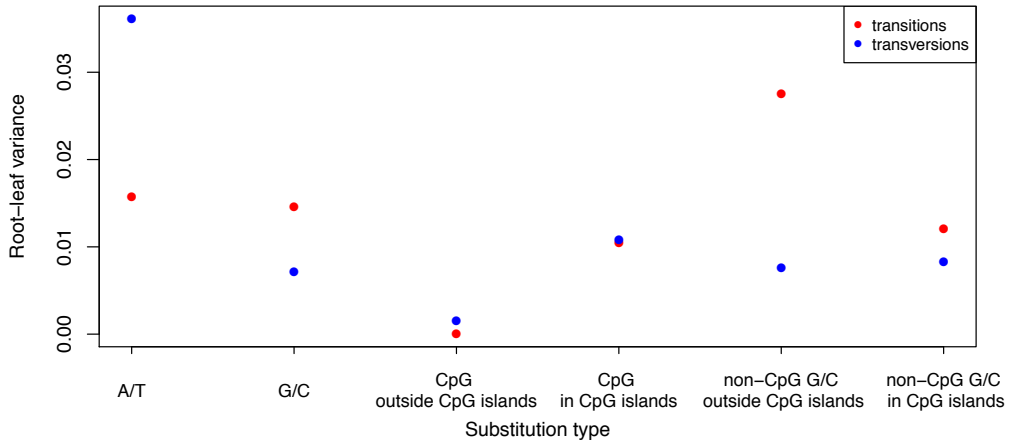
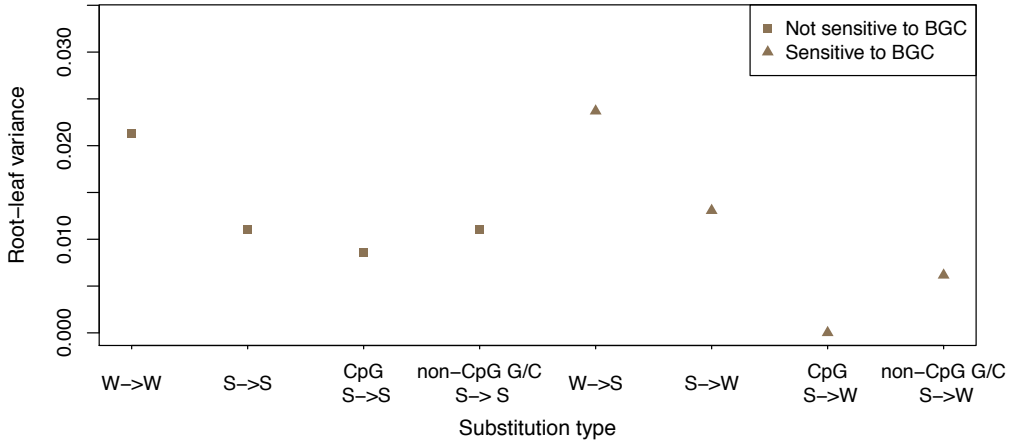


Figure S8: Effect of biased gene conversion across lineages estimated for different datasets. For each substitution type (strong (S; G/C) and weak (W; A/T)) and each ancestral context shown on the x-axis, we estimated the total branch length from the root to each terminal leaf as the inferred number of substitutions per site, in (a) Multiz and (b) EPO dataset. We then computed the variance in the normalized root to leaf distance across five primates (human, chimpanzee, orangutan, rhesus macaque and baboon). This figure differs from Figure 2B, in that it uses fewer species in the Multiz dataset in order to match the set of species (hominoids and OWMs) available in the EPO dataset. R code to replicate this figure is available at:
https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS8.R

(a) Multiz



(b) EPO

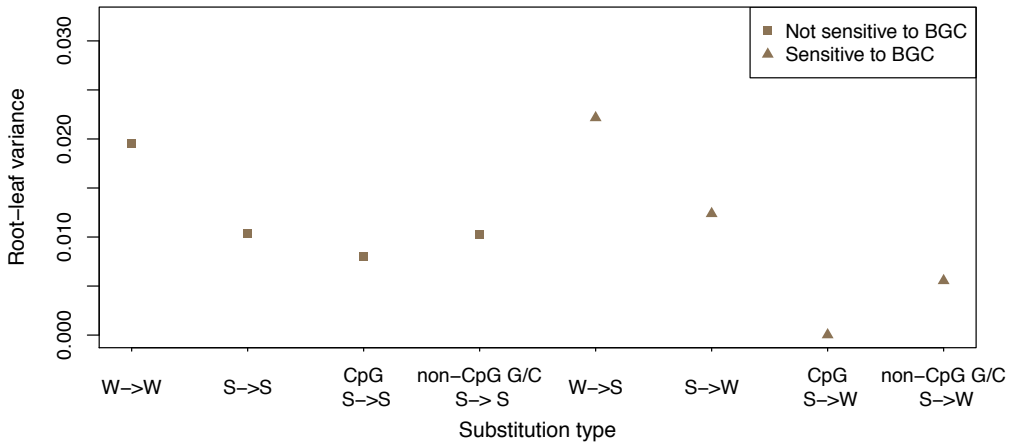


Figure S9: Comparison of substitution rates in human and chimpanzee using Phylofit. For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio of the substitution rate in chimpanzee to the substitution rate in human is shown as the title of each subfigure. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS9.R

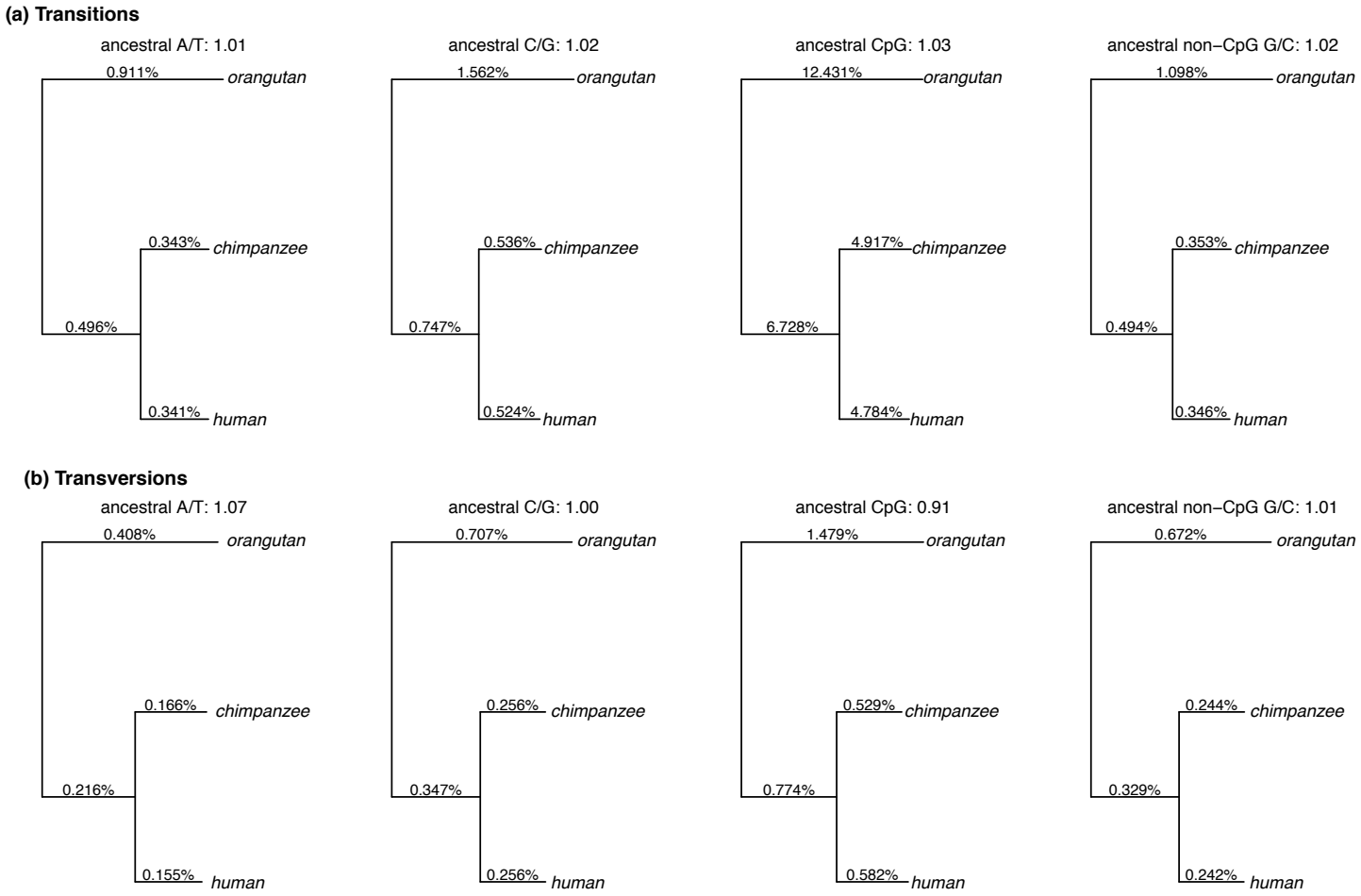


Figure S10: Comparison of substitution rates in human and gorilla using Phylofit. For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of the substitution rate in gorilla to the substitution rate in human is shown as the title of each subfigure. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS10.R

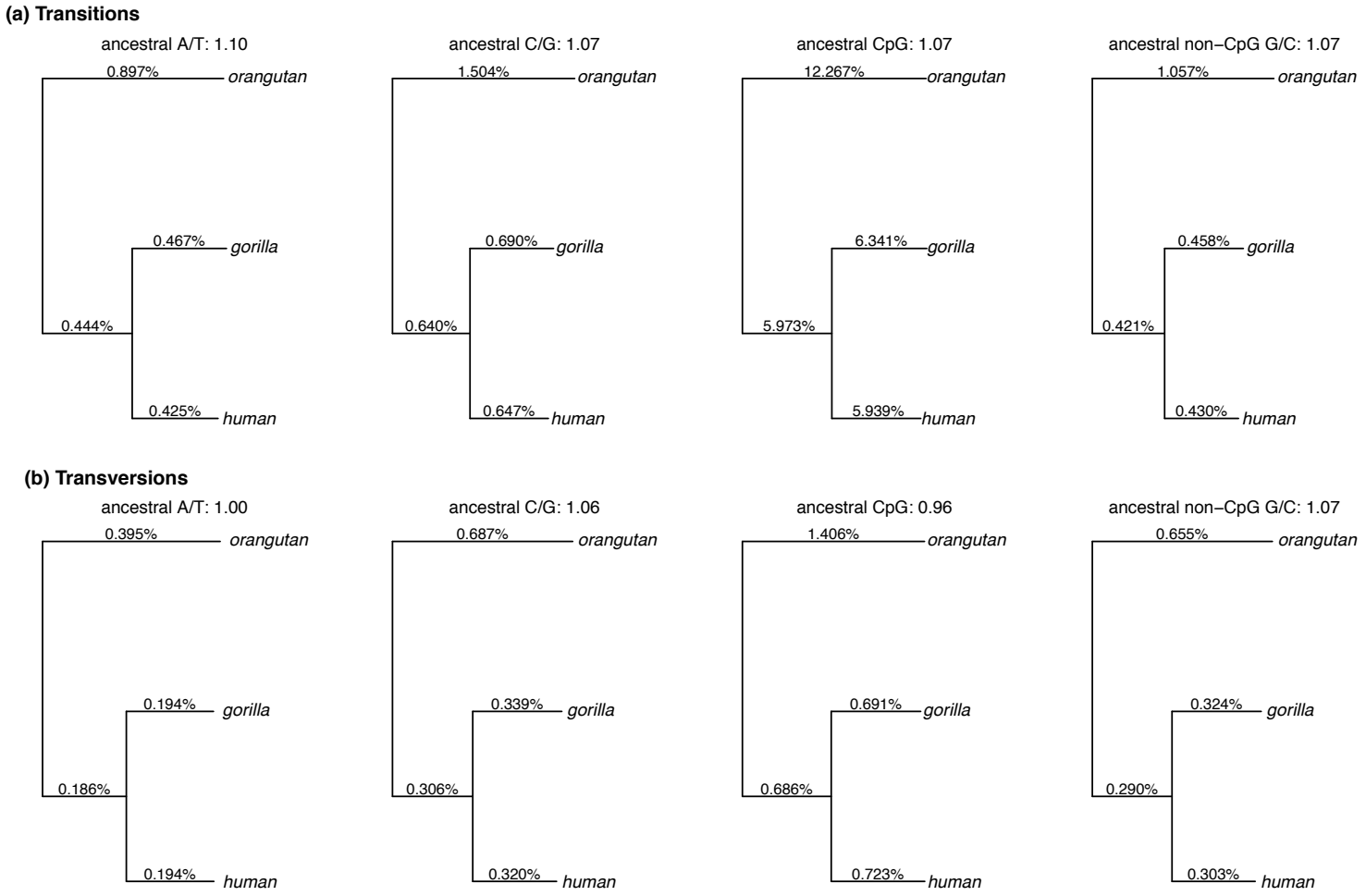


Figure S11: Comparison of substitution rates in human and chimpanzee using the maximum likelihood approach. For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio of the substitution rate in chimpanzee to the substitution rate in human is shown as the title of each subfigure. The maximum likelihood method does not estimate rates for all ancestral G/C sites (i.e., it only reports CpG and non-CpG G/C rates separately) and hence we do not report results for this context. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS11.R

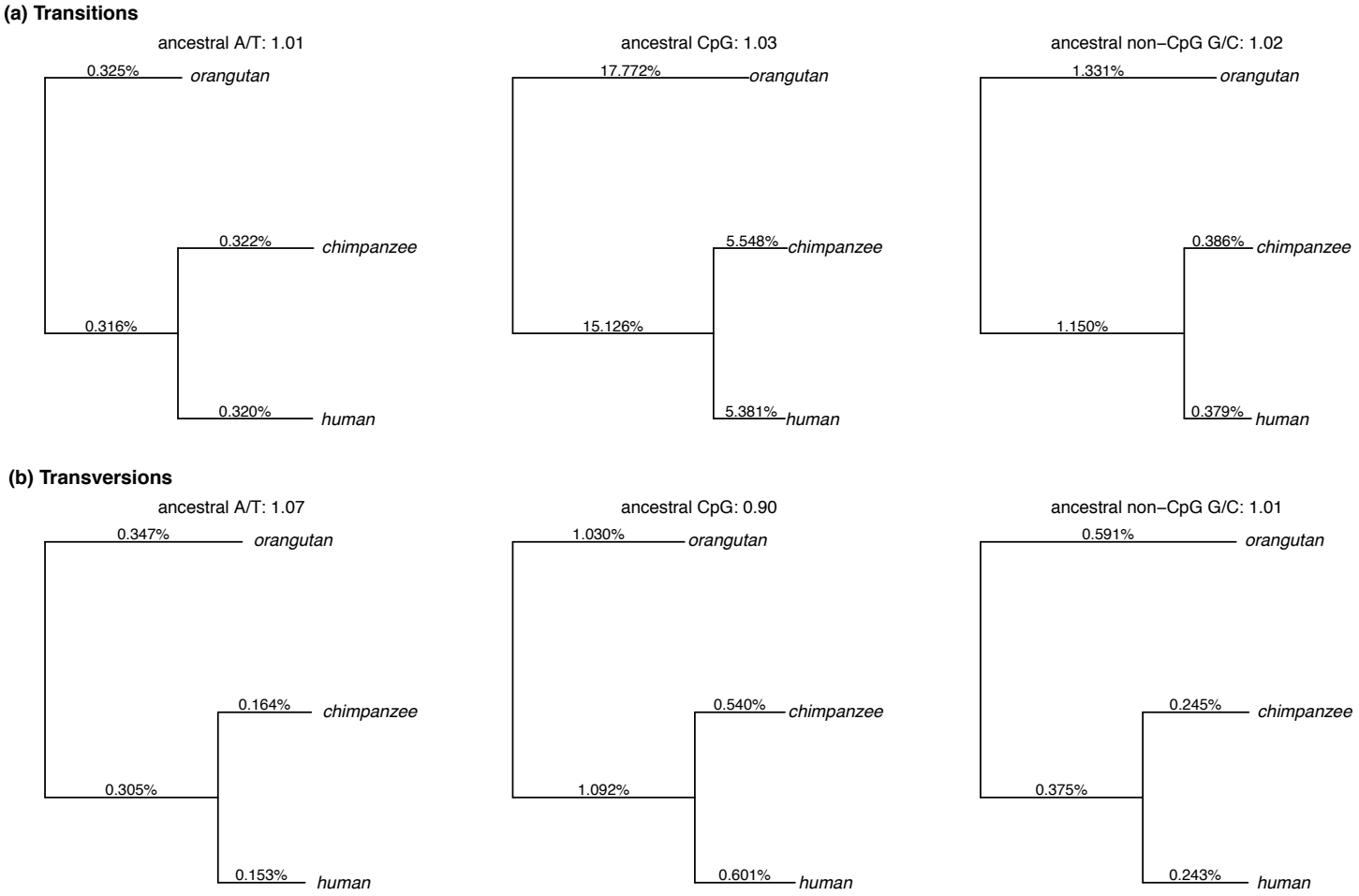


Figure S12: Comparison of substitution rates in human and gorilla using the maximum likelihood approach. For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of the substitution rate in gorilla to the substitution rate in human is shown as the title of each subfigure. The maximum likelihood method does not estimate the rates for all ancestral G/C sites (i.e., it only reports CpG and non-CpG G/C rates separately) and hence we do not report results for this context. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS12.R

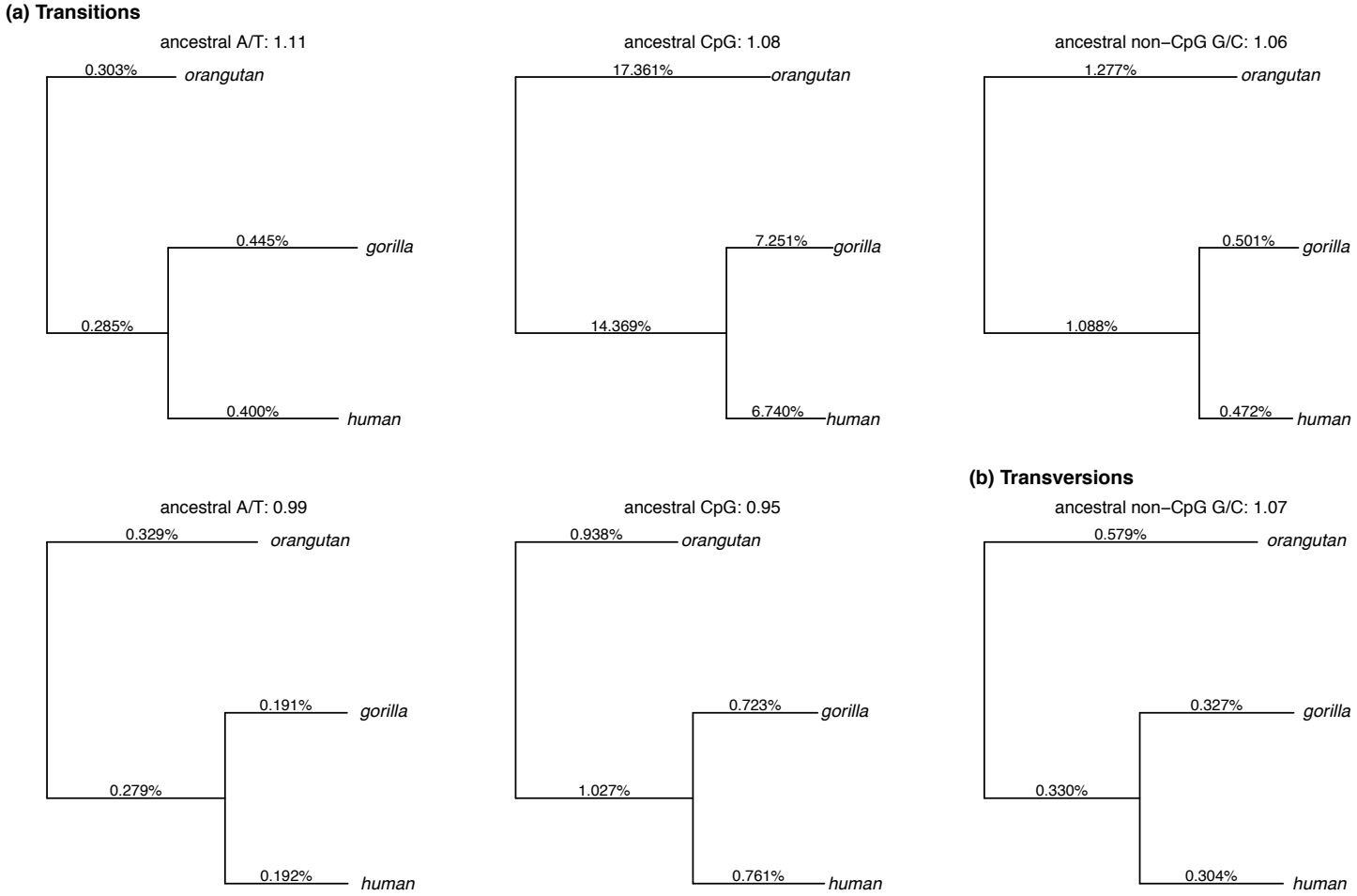


Figure S13: Mutation spectrum across primates. We estimated the number of substitutions along each lineage for each mutation type. We then normalized the number of substitutions of a given type to the number of transitions from ancestrally CpG sites that occurred on that lineage. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/FigureS13.R

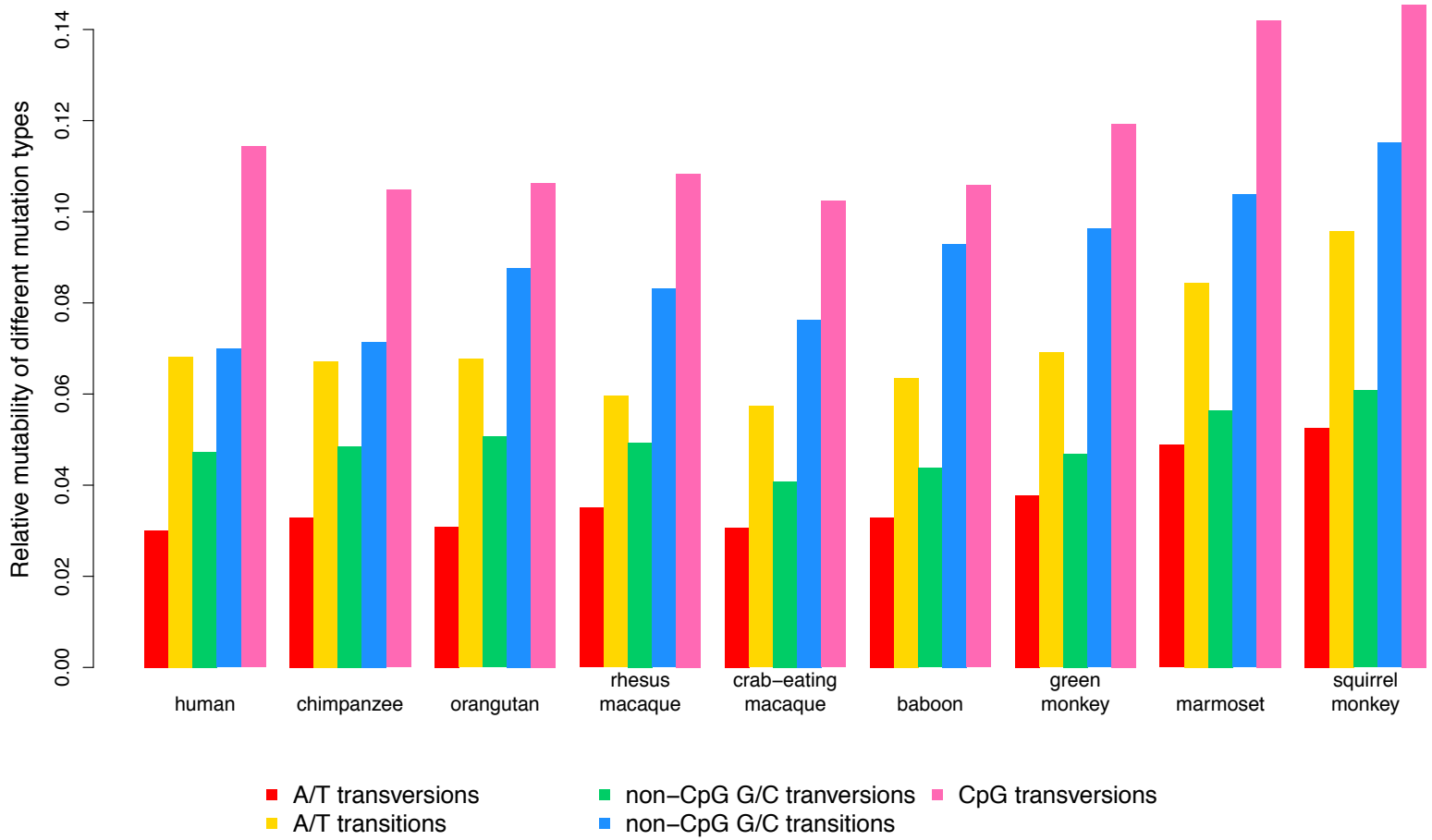


Table S1. Online source of annotation for transposable elements, coding exons, CpG Islands (CGI), and conserved sites.

Assembly	Annotation			Dataset
	Transposable elements	Coding exons	CGI	
hg19	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/cpgIslandExt.txt.gz	Multiz, high coverage
hg38	http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/cpgIslandExt.txt.gz	EPO
panTro4	http://hgdownload.soe.ucsc.edu/goldenPath/panTro4/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/panTro4/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/panTro4/database/cpgIslandExt.txt.gz	EPO, Multiz, high coverage
gorGor3	http://hgdownload.soe.ucsc.edu/goldenPath/gorGor3/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/gorGor3/database/ensGene.txt.gz	No annotation available	EPO, Multiz, high coverage
ponAbe2	http://hgdownload.soe.ucsc.edu/goldenPath/ponAbe2/database/chr*_rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/ponAbe2/database/refGene.txt.gz	No annotation available	EPO, Multiz, high coverage
nomLeu3	http://hgdownload.soe.ucsc.edu/goldenPath/nomLeu3/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/nomLeu3/database/genscan.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/nomLeu3/database/cpgIslandExt.txt.gz	Multiz
rheMac2	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac2/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac2/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac2/database/cpgIslandExt.txt.gz	EPO
rheMac3	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac3/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac3/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/rheMac3/database/cpgIslandExt.txt.gz	Multiz
macFas5	No annotation available	No annotation available	No annotation available	Multiz
papHam1	http://hgdownload.soe.ucsc.edu/goldenPath/papHam1/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/papHam1/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/papHam1/database/cpgIslandExt.txt.gz	Multiz
papAnu2	http://hgdownload.soe.ucsc.edu/goldenPath/papAnu2/database/rmsk.txt	http://hgdownload.soe.ucsc.edu/goldenPath/papAnu2/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/papAnu2/database/c	EPO

	.gz		pgIslandExt.txt.gz	
chlSab1	No annotation available	No annotation available	No annotation available	Multiz
calJac3	http://hgdownload.soe.ucsc.edu/goldenPath/calJac3/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/calJac3/database/refGene.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/calJac3/database/cpgIslandExt.txt.gz	EPO, Multiz
saiBol1	http://hgdownload.soe.ucsc.edu/goldenPath/saiBol1/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/saiBol1/database/genscan.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/saiBol1/database/cpgIslandExt.txt.gz	Multiz
otoGar3	http://hgdownload.soe.ucsc.edu/goldenPath/otoGar3/database/rmsk.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/otoGar3/database/genscan.txt.gz	http://hgdownload.soe.ucsc.edu/goldenPath/otoGar3/database/cpgIslandExt.txt.gz	Multiz
Conserved sites				
hg19	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/phastConsElements46wayPrimates.txt.gz			EPO, Multiz, high coverage

Table S2: Estimate of various life history traits for different primate species

Species	Common name	Gestation time (in days)[a]	SECL (in days)[b]	Onset of puberty in males (in years)[c]	Ratio of male to female generation time	Mean sex-averaged generation time (in years)
Homo sapiens	Human	280	16 (30)	13.5 (31)	1.1 (32-34)	29 (32)
Pan troglodytes	Chimp	229	14 (35)	8.5 (36)	0.96 (37)	25 (37)
Gorilla gorilla	Gorilla	256	--	7 (38)	1.1 (37)	19 (37)
Pongo abelii	Orangutan	249	--	6.5[+] (39)	--	27[§] (40)
Macaca fascicularis	Crab-eating macaque	165	10.2 (41)	3.5 (42)	*	11[+] (43)
Macaca mulatta	Rhesus macaque	165	10.5 (44)	3.5 (42)	*	12 (43)
Papio anubis	Baboon	171[+]	11 (45)	5.4[+] (46)	*	11 (47)
Cercopithecus aethiops	Green Monkey	132[+]	10.2 (48)	5 (42)	--	11[+] (49)
Saimiri sciureus	Squirrel Monkey	161	10.2 (48)	3 (50)	*	9[§] (51)
Callithrix jacchus	Marmoset	144	10 (52)	0.9 (53)	*	6 (54)

Note: -- = not available. [§] only female generation was available. [+] inferred from a closely related species.

[a] source: AnAge: The animal ageing and longevity database, build 13.

[b] source: (55) and references within.

[c] main source: (50) and other papers listed.

[*] not available so assumed to be 1.0 when modeling yearly mutation rates in these species.

Table S3: Autosomal substitution rates on the human lineage for different time depths and using different filters.

Sequence	Selective constraint	H-HC	H-HO	H-HM
Whole genome	-	0.56%	1.46%	2.51%
CET	putatively non-neutral	0.51%	1.26%	2.23%
Whole genome - CET	putatively neutral	0.58%	1.52%	2.65%
AR	putatively neutral	0.58%	1.50%	2.62%

Note: CET = conserved elements, exons and transposable elements, AR = Ancestral repeats.
To identify putatively neutral AR, we considered all transposable elements (i.e. LINE, SINE, LTR or DNA elements) that are shared between human (hg19) and rhesus macaque (rheMac3) genomes based on UCSC Table Browser. Following Ananda et al. (56), we excluded L1PA1-A7, L1HS, and AluY as these were inserted in the human genome after to the human-macaque divergence and MER121 that have been shown to be under strong selection (57).

Table S4: Correlation in life history traits across primates.

Trait	SECL	Generation time	Onset of Puberty	G-P
SECL	1	0.90**	0.91**	0.71*
Generation time (G)	--	1	0.89***	0.92***
Onset of puberty (P)	--	--	1	0.74*
G-P	--	--	--	1

Note: Estimates based on Spearman's rank correlation corrected for ties.
Significance codes: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

References:

1. Karolchik D*, et al.* (2014) The UCSC genome browser database: 2014 update. *Nucleic acids research* 42(D1):D764-D770.
2. Paten B, Herrero J, Beal K, Fitzgerald S, & Birney E (2008) Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome research* 18(11):1814-1828.
3. Earl D*, et al.* (2014) Alignathon: a competitive assessment of whole-genome alignment methods. *Genome research* 24(12):2077-2089.
4. Venn O*, et al.* (2014) Strong male bias drives germline mutation in chimpanzees. *Science* 344(6189):1272-1275.
5. Prado-Martinez J*, et al.* (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471-475.
6. Locke DP*, et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469(7331):529-533.
7. Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30(20):2843-2851.
8. Siepel A*, et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* 15(8):1034-1050.
9. Murphy WJ*, et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294(5550):2348-2351.
10. Meunier J, Khelifi A, Navratil V, & Duret L (2005) Homology-dependent methylation in primate repetitive DNA. *proceedings of the national Academy of Sciences of the United States of America* 102(15):5471-5476.
11. Smit A, Hubley R, & Green P (2004) RepeatMasker Open-3.0. 2004. *Seattle (WA): Institute for Systems Biology*.
12. Takai D & Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the national academy of sciences* 99(6):3740-3745.
13. Molaro A*, et al.* (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146(6):1029-1041.
14. Gardiner-Garden M & Frommer M (1987) CpG islands in vertebrate genomes. *Journal of molecular biology* 196(2):261-282.
15. Siepel A & Haussler D (2004) Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Molecular Biology and Evolution* 21(3):468-488.
16. Duret L & Arndt PF (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* 4(5):e1000071.
17. Mailund T, Munch K, & Schierup MH (2014) Lineage sorting in apes. *Annual review of genetics* 48:519-535.
18. Jombart T & Dray S (2013) adephylo: exploratory analyses for the phylogenetic comparative method.
19. Hwang DG & Green P (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* 101(39):13994-14001.

20. Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*:1-15.
21. Paradis E (2011) *Analysis of Phylogenetics and Evolution with R* (Springer Science & Business Media).
22. Amster G & Sella G (2016) Life history effects on the molecular clock of autosomes and sex chromosomes. *Proceedings of the National Academy of Sciences* 113(6):1588-1593.
23. Ségurel L, Wyman MJ, & Przeworski M (2014) Determinants of mutation rate variation in the human germline. *Annual review of genomics and human genetics* 15:47-70.
24. Kong A*, et al.* (2012) Rate of de novo mutations and the importance of father/'s age to disease risk. *Nature* 488(7412):471-475.
25. Gao Z, Wyman MJ, Sella G, & Przeworski M (2016) Interpreting the dependence of mutation rates on age and time. *PLoS biology* 14(1):e1002355.
26. Wall JD (2003) Estimating ancestral population sizes and divergence times. *Genetics* 163(1):395-404.
27. Meyer M*, et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-226.
28. Prüfer K*, et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-49.
29. Kong A*, et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467(7319):1099-1103.
30. Heller C & Clermont Y (1963) Kinetics of the germinal epithelium in man. *Recent progress in hormone research* 20:545-575.
31. Marshall WA & Tanner JM (1970) Variations in the pattern of pubertal changes in boys. *Archives of disease in childhood* 45(239):13-23.
32. Fenner JN (2005) Cross‐cultural estimation of the human generation interval for use in genetics‐based population divergence studies. *American journal of physical anthropology* 128(2):415-423.
33. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, & Stefánsson K (2003) A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *The American Journal of Human Genetics* 72(6):1370-1388.
34. Matsumura S & Forster P (2008) Generation time and effective population size in Polar Eskimos. *Proceedings of the Royal Society of London B: Biological Sciences* 275(1642):1501-1508.
35. Smithwick E, Young L, & Gould K (1996) Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. *Tissue and Cell* 28(3):357-366.
36. Behringer V, Deschner T, Deimel C, Stevens J, & Hohmann G (2014) Age-related changes in urinary testosterone levels suggest differences in puberty onset and divergent life history strategies in bonobos and chimpanzees. *Hormones and behavior* 66(3):525-533.
37. Langergraber KE*, et al.* (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences* 109(39):15716-15721.

38. Harcourt AH, Fossey D, Stewart KJ, & Watts DP (1979) Reproduction in wild gorillas and some comparisons with chimpanzees. *Journal of reproduction and fertility. Supplement*:59-70.
39. Dixson A, Knight J, Moore H, & Carman M (1982) Observations on sexual development in male Orang‐utans. *International Zoo Yearbook* 22(1):222-227.
40. Wich SA, de Vries, H., Ancrenaz, M., Perkins, L., Shumaker, R. W., Suzuki, A., and van Schaik, C. P. (2009) Orangutan life history variation. In: Wich, Serge A (2009) *Orangutans: geographic variation in behavioral ecology and conservation* (Oxford University Press).
41. Aslam H*, et al.* (1999) The cycle duration of the seminiferous epithelium remains unaltered during GnRH antagonist-induced testicular involution in rats and monkeys. *Journal of endocrinology* 161(2):281-288.
42. Bercovitch FB (2000) Behavioral ecology and socioendocrinology of reproductive maturation in cercopithecine monkeys. *Old world monkeys*:298-320.
43. Molur S & Organisation ZO (2003) *Status of South Asian Primates: Conservation Assessment and Management Plan (CAMP), Workshop Report, 2003* (Zoo Outreach Organisation and Conservation Breeding Specialist Group, South Asia in collaboration with Wildlife Information & Liaison Development Society).
44. De Rooij D, van Alphen M, & van de Kant H (1986) Duration of the cycle of the seminiferous epithelium and its stages in the rhesus monkey (Macaca mulatta). *Biology of reproduction* 35(3):587-591.
45. Chowdhury A & Steinberger E (1976) A study of germ cell morphology and duration of spermatogenic cycle in the baboon, Papio anubis. *The Anatomical record* 185(2):155-169.
46. Onyango PO, Gesquiere LR, Altmann J, & Alberts SC (2013) Puberty and dispersal in a wild primate population. *Hormones and behavior* 64(2):240-249.
47. Altmann J, Gesquiere L, Galbany J, Onyango PO, & Alberts SC (2010) Life history context of reproductive aging in a wild primate model. *Annals of the New York Academy of Sciences* 1204(1):127-138.
48. Barr A (1973) Timing of spermatogenesis in four nonhuman primate species. *Fertility and sterility* 24(5):381-389.
49. Isbell LA, Young TP, Jaffe KE, Carlson AA, & Chancellor RL (2009) Demography and life histories of sympatric patas monkeys, Erythrocebus patas, and vervets, Cercopithecus aethiops, in Laikipia, Kenya. *International journal of primatology* 30(1):103-124.
50. Dixson AF (2009) *Sexual selection and the origins of human mating systems* (Oxford University Press).
51. Abee CR, Mansfield K, Tardif SD, & Morris T (2012) *Nonhuman Primates in Biomedical Research: biology and management* (Academic Press).
52. Millar MR, Sharpe RM, Weinbauer GF, Fraser HM, & Saunders PT (2000) Marmoset spermatogenesis: organizational similarities to the human. *International journal of andrology* 23(5):266-277.
53. Abbott DH, Barnett DK, Colman RJ, Yamamoto ME, & Schultz-Darken NJ (2003) Aspects of common marmoset basic biology and life history important for biomedical research. *Comparative medicine* 53(4):339-350.

54. Gage TB (1998) The comparative demography of primates: with some comments on the evolution of life histories. *Annual Review of Anthropology*:197-221.
55. Ramm SA & Stockley P (2009) Sperm competition and sperm length influence the rate of mammalian spermatogenesis. *Biology letters*:rsbl20090635.
56. Ananda G, Chiaromonte F, & Makova KD (2011) A genome-wide view of mutation rate co-variation using multivariate analyses. *Genome biology* 12(3):R27.
57. Kamal M, Xie X, & Lander ES (2006) A large family of ancient repeat elements in the human genome is under strong selection. *Proceedings of the National Academy of Sciences of the United States of America* 103(8):2740-2745.