Variation in the molecular clock of primates

Priya Moorjani[1,2*,+], Carlos Eduardo G. Amorim[1*], Peter F. Arndt[3], Molly Przeworski[1,4,+]

[1] Department of Biological Sciences, Columbia University, New York, NY
[2] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA
[3] Max Planck Institute for Molecular Genetics, Berlin, Germany
[4] Dept. of Systems Biology, Columbia University, New York, NY
[*] Contributed equally

[+]To whom correspondence should be addressed: Priya Moorjani (pm2730@columbia.edu) and Molly Przeworski (mp3284@columbia.edu)

Abstract

Events in primate evolution are often dated by assuming a "molecular clock", i.e., a constant rate of substitution per unit time, but the validity of this assumption remains unclear. Among mammals, it is well known that there exists substantial variation in yearly substitution rates. Such variation is to be expected from differences in life-history traits, suggesting that it should also be found among primates. Motivated by these considerations, we analyze whole genomes from ten primate species, including Old World Monkeys (OWMs), New World Monkeys (NWMs) and apes, focusing on putatively neutral autosomal sites and controlling for possible effects of biased gene conversion and methylation at CpG sites. We find that substitution rates are ~65% higher in lineages leading from the hominoid-NWM ancestor to NWMs than to apes. Within apes, rates are ~2% higher in chimpanzees and ~7% higher in the gorilla than in humans. Substitution types subject to biased gene conversion show no more variation among species than those not subject to it. Not all mutation types behave similarly, however: in particular, transitions at CpG sites exhibit a more clock-like behavior than do other types, presumably due to their non-replicative origin. Thus, not only the total rate, but also the mutational spectrum varies among primates. This finding suggests that events in primate evolution are most reliably dated using CpG transitions. Taking this approach, we estimate that the average time to the most recent common ancestor of human and chimpanzee is 12.1 million years and their split time 7.9 million years.

Significance statement

Much of our understanding of the chronology of human evolution relies on the "molecular clock", i.e., a constant rate of substitutions per unit time. To evaluate the validity of this assumption, we analyze whole genome sequences from ten primate species. We find that there is substantial variation in the molecular clock between apes and monkeys, and rates even differ within hominoids. Importantly, not all mutation types behave similarly: notably, transitions at CpG sites exhibit a more clock-like behavior than other substitutions, presumably due to their non-replicative origin. Thus, the mutation spectra, and not just the overall substitution rates, are changing across primates. This finding further suggests that events in primate evolution are most reliably dated using CpG transitions.

Introduction

Germline mutations are the ultimate source of genetic differences among individuals and species. They are thought to arise from a combination of errors in DNA replication (e.g., the chance misincorporation of a base pair) or damage that is unrepaired by the time of replication (e.g., the spontaneous deamination of methylated CpG sites) (1). If mutations are neutral (i.e., do not affect fitness), then the rate at which they arise will be equal to the substitution rate (2). A key consequence is that if mutation rates remain constant over time, substitution rates should likewise be constant.

This assumption of constancy of substitution rates plays a fundamental role in evolutionary genetics, by providing a molecular clock by which to date events inferred from genetic data (3). Notably, important events in human evolution for which there is no fossil record (e.g., when humans and chimpanzees split, or when anatomically modern humans first left the African subcontinent) are dated using a mutation rate obtained from contemporary pedigrees or phylogenetic analysis, assuming the per year rate has remained unchanged for millions of years (4, 5).

Yet we know from studies of mammalian phylogenies, as well as of other taxa, that there can be substantial variation in substitution rates per unit time (6-8). In particular, there is the well known "generation time effect" on substitution rates, which suggests that species with shorter generation time (i.e. mean age of reproduction) have higher mutation rates (9). For instance, mice have a generation time on the order of months (~10-12 months) compared to ~29 years in humans, and a two- to three-fold higher substitution rate per year (9). More generally, in a survey of 32 mammalian species, the strongest predictor of substitution rate variation was the generation time (6).

A generation time effect has also been suggested in humans as the basis for a possible slowdown in the yearly mutation rate towards the present. This conjecture is motivated by the observation that the yearly mutation rate estimated from substitutions accumulated between humans and chimpanzees over millions of years (~10^{-9} per base pair per year (10, 11)) is more than two-fold higher than the yearly mutation rate estimated from *de novo* mutation events in trios (~0.4×10^{-9} per base pair per year (12, 13)). One possibility is that the generation time has increased towards the present, and led to a decrease in the yearly mutation rate (14).

Whether the association between generation time and substitution rates is causal remains unclear, however, correlated traits such as metabolic rate (15), body size (16) and sperm competition (17), may also affect substitution rates. For instance, the metabolic rate hypothesis posits that species with higher basal metabolic rates are subject to higher rates of oxidative stress and hence have a higher mutation rate (15). Body mass has been shown to be negatively correlated to substitution rates, such that smaller animals tend to have higher substitution rates (15). Sexual selection on mating systems may also affect substitution rates, more intense sperm competition leads to selection for higher sperm counts, leading to more cell divisions per unit time during spermatogenesis and a higher male mutation rate (17).

3

That said, an effect of life history traits such as generation time on the yearly mutation rate is expected based on first principles, given our understanding of oogenesis and spermatogenesis (18, 19). In mammals, oogonial divisions are completed by the birth of the future mother, whereas the spermatogonial stem cells continue to divide post-puberty (18). Thus, the total number of replication-driven mutations inherited by a diploid offspring accrues in a piecewise linear manner with parental age, with the number depending on the number of cell divisions in each developmental stage as well as the per cell division mutation rates (1, 19). These considerations indicate that changes in generation time, onset of puberty and rate of spermatogenesis should all influence yearly mutation rates (1, 19).

Importantly then, primates are well known to differ with regard to most of these traits. In addition to huge variation in body size and metabolic rates, generation time varies almost ten fold, with the shortest generation time observed in prosimians (~3 years in galago and mouse lemurs (20)) and the longest generation time observed in humans (~29 years (21)). Species also differ in the strength of sperm competition and rates of spermatogenesis: monkeys have a shorter spermatogenetic division and thus consequently produce more sperm per unit time than do apes (22). Thus, even if the per cell division mutation rate remained constant, we should expect differences in yearly mutation rates among species.

While the factors discussed thus far apply to all sites, variation in substitution rates among species also depends on the type of mutation and the genomic context (i.e., flanking sequence) in which it occurs (7). For example, in mammals, CpG transitions show the least amount of variation in substitution rates among species (7). A plausible explanation is the source of mutations: transitions at methylated CpG sites are thought to occur primarily through spontaneous deamination; if they arise at a constant rate and their repair is inefficient relative to the cell cycle length, as is thought to be the case, then their mutation rate should depend largely on absolute time rather than on the number of cell divisions (23-25).

In addition, even substitutions that are neutral in their effects on fitness may vary in their rate of accumulation among lineages because of biased gene conversion (BGC), the bias towards strong (S: G or C) rather than weak (W: A or T) bases that occurs in the repair of double strand breaks (26). This phenomenon leads to the increased probability of fixation of S alleles (and loss of W alleles) in regions of higher recombination, and can therefore change substitution rates relative to mutation rates (26, 27). The strength of BGC is a function of the degree of bias, the local recombination rate and the effective population size of the species (26). The latter varies by three to four fold among primates (28), and the fine-scale recombination landscape is also likely to differ substantially across species (29, 30).

Empirically, the extent to which substitution rates vary among primate lineages remains unclear. Kim et al. (2006) compared two hominoids (human and chimpanzee) and two OWMs (baboon and rhesus macaque); assuming that the average divergence time of the two pairs of species is identical, they reported that substitution rates at transitions at non-CpG sites differ by approximately 31% between hominoids and OWM, whereas rates of

CpG transitions are almost identical (31). In turn, Elango et al. (2006) found that the human branch is ~2% shorter than chimpanzees (considering the rates from the human-chimpanzee ancestor), and ~11% shorter than gorilla (considering rates from the human-gorilla ancestor) (32). While these comparisons raise the possibility that substitution rates are evolving across primates, they are based on limited data, make strong assumptions about divergence times, and rely on parsimony-based approaches that may underestimate substitution rates for divergent species, notably at CpG sites (33). We therefore revisit these questions using whole genome sequence alignments of ten primates, allowing for variable substitution rates along different lineages and explicitly modeling the context dependency of CpG substitutions.

Materials and Methods

Data sets and filtering. We used the following datasets for our analysis: (a) A 12-primate whole genome sequence alignment, with mouse as an outgroup, which is part of a 100-way mammalian phylogeny, mapped using Multiz (34) (referred to as the Multiz dataset). (b) A seven primate whole genome alignment, mapped using the Enredo-Pecan-Ortheus (EPO) pipeline (35) (referred to the EPO dataset); and (c) High coverage genomes for a human (of European descent) that we sequenced (Note S1), a chimpanzee (Ind-D from (36)) and a gorilla (Delphi from (37); data kindly provided by Tomas Marques-Bonet, Institut Biologia Evolutiva (Universitat Pompeu Fabra/CSIC)) (referred to as the high coverage hominoid dataset). For the EPO dataset (b), we removed duplications using the mafDuplicateFilter from mafTools package (38). This software identifies any duplicated region in the alignment block and only retains the sequence with the highest similarity to the consensus sequence. In (c), the genomes were mapped to the orangutan reference genome (ponAbe2) (39), which should be equidistant to humans and extant African great apes (assuming no variation in substitution rates), using bwa-mem (40) 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. Single Nucleotide polymorphisms (SNP) were called using samtools mpileup (version: 0.1.18-dev) (41) 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, using the orangutan reference genome as an 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 minimize the effects of selection acting on the base pairs considered, we limited our analyses to the non-coding, non-conserved and non-repetitive regions of the genome. Table S1 includes the source of all annotations used. For each primate species, we excluded sites with the following annotations:

5

(a) Conserved elements annotated using phastCons (42) based on the multiple alignments of 46 primates (43). 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 (44), 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 (45).

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 (46). As an illustration, comparison of sperm methylation profiles in humans from (47) 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 (Supplementary 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) (48).

Estimating substitution rates. We used Phylofit (49) to estimate autosomal substitutions from the three datasets described above. To access the robustness of the estimates from Phylofit, we also used an alternative maximum likelihood based approach from (50) 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 dataset, the EPO dataset or that of hominoids. 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 chimpanzees/gorillas and gibbons/orangutans, we considered only one of the two lineages in a given analysis (51).

Phylofit (49) analysis was performed with the expectation maximization algorithm (option -E) with medium precision for convergence. 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. For both internal and external branches, Phylofit outputs both the overall branch lengths, accounting for recurrent substitutions at a site, and posterior expected counts of the number of positions at which the descendant base differs from the ancestral base (option -Z).

6

We used the posterior counts 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 context, we estimated the lineage-specific divergence from any internal node to the terminal node X such that $X_{A1->A2}$, is the number of positions at which the ancestral allele A_1 (at the internal node) is estimated to have substituted to allele A_2 (at the terminal node) and divided by the total count of ancestral alleles A_1 on the lineage X, assuming single-step mutations from A_1 to A_2, thereby implicitly 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). We observed that the estimates were quite unstable, and in particular that a subset of the runs had substantially lower likelihoods. We interpreted this result as indicating that the method sometimes returns values for a local peak in the likelihood surface. To circumvent this problem, we therefore took the estimate for the run with the highest likelihood. Additionally, we also used the maximum likelihood based approach from (50). This approach uses a probabilistic model for sequence evolution and assumes that all nucleotide substitutions except those occurring in a CpG context evolve independently. This leads to 6 parameters in a reverse complement symmetric analysis or 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 3 or 6 additional parameters in the reverse complement symmetric setting or non-reverse complement symmetric setting, respectively. Because the identity of left neighbor or the right neighbor influences substitution processes (e.g. CpG -> TpG or CpG -> CpA), 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 nucleotide contents in the orangutan genome for normalization.

Assessing the significance of branch length differences in pairwise comparisons. To test if the branch lengths estimated by Phylofit differ between two species (species$_1$ and species$_2$), we used a likelihood ratio test where the null model (H$_o$) was that the number of substitutions on the branch leading to species$_1$ (n_1) and branch leading to species$_2$ (n_2) from the common ancestor are equal, so the proportion (p) $= n_1/n = 0.5$, where $n = n_1 + n_2$. In the alternative model (H$_A$), the number of substitutions on the branch leading to species$_1$ was not equal to the number of substitutions on the branch to species$_2$, i.e p $\neq 0.5$. Thus, the likelihood ratio statistic

7

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

should be approximately $\chi^2(df = 1)$.

Estimating the root-leaf variance. For each substitution type, we constructed a phylogenetic tree using the lineage-specific substitution rates estimated by Phylofit from the Multiz and EPO datasets. We computed the root-leaf distance using the *R* package *adephylo* (52). Following (7), we considered the variance in the root to leaf distance after normalizing by the mean distance.

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 (1985) (53) implemented in the *R* package *ape* (54). We used the CpG transition rates to define our baseline 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 (25). In this model, the mutation rate per year is given by:

$$\mu_y = \frac{\mu_4^m c^m}{2} + \frac{A^*}{2G}$$

where $A^* = 2\mu_0 + 2\mu_1 d_1 + \mu_2^f d_2^f + \mu_2^m d_2^m + \mu_3^m d_3^m - \mu_4^m \left(c^m P + c^m t_{sg} - d_{sg}\right)$. Here, μ_i^s, d_i^s is the mutation rate and number of cell divisions in the ith stage of development (i = 0: first post-zygotic division, 1: second post-zygotic division to sex differentiation, 3: sex differentiation to birth, 4: birth to puberty and 5: puberty to reproduction) in sex s (male or female) respectively, t_{sg} is the duration of spermatogenesis (in years), d_{sg} is the number of spermatogonial stem cell divisions required to complete spermatogenesis, $c^m = \frac{365}{SECL}$ 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)), P is the age of puberty and G is the mean age of reproduction (assumed to be the same in males and females).

Following (25), we assumed $d_1 = 15$, $d_2^f = 15$, $d_2^m = 21$, $d_3^m = 0$, $d_{sg} = 4$, $t_{sg} = 0.2$ and $\mu_2 = \mu_3 = \mu_4 = \mu$ and $\mu_o = \mu_1 = 4\mu$, to allow for a higher mutation rate in the first two post-zygotic divisions (55). Parameter values for life-history traits used for different species are shown in Table S2.

Estimating divergence and split times among apes from CpG substitutions. We estimated the divergence time between human-chimpanzee and human-gorilla using substitutions involving transitions at CpG sites (outside CGI), by:

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

8

where $X_{CG \to TG/CA}$ is the expected number of transitions estimated by Phylofit to have 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 $\mu_{CG \to TG/CA}$ is the estimated per year mutation rate for CpG transitions. The value of $\mu_{CG \to TG/CA}$ ($=3.9 \times 10^{-9}$ per base pair per year) was obtained by dividing the per generation mutation rate at CpG transitions ($= 1.12 \times 10^{-7}$ per base pair per generation) in (12) by the mean parental age in that study (28.4 years), which is appropriate if the number of CpG transitions is considered linear with age.

As $t_{divergence} = t_{split} + t_{coalescent}$ and assuming a panmictic, constant size population, $t_{coalescent} = 2N_a G$,

$$t_{split} = \frac{X_{CG \to TG/CA}}{\mu_{CG \to TG/CA}} - 2N_a G$$

where N_a is effective population size of the ancestral population. We assumed $N_a = 5N_h$ (37, 56) and $N_h = \pi_{CG \to TG/CA} / 4\mu_{CG \to TG/CA}$, where N_h is the effective population size in humans and $\pi_{CG \to TG/CA}$ is the average diversity level observed at transitions at CpG sites across 13 contemporary human populations (57, 58).

Results

We first estimate the number of autosomal substitutions on 10 primate lineages by applying Phylofit (49) to the Multiz sequence alignment (excluding gorilla and gibbon due to concerns about incomplete lineage sorting; see Methods). This method allows us to estimate branch lengths, accounting for uncertainty in the ancestral reconstruction, recurrent substitutions at a site, and allowing for context dependent effects of neighboring nucleotides at CpG dinucleotides (49).

To focus on putatively neutral sites in the genome, in which substitutions more faithfully reflect mutation rates and patterns, we exclude conserved elements, coding exons and transposable elements (referred to as CET in what follows, see Methods). After filtering CET sites and removing missing data, we obtain ~562 Mb of whole genome sequence alignment across 10 primates. In these data, the total substitution rates vary markedly across the 10 primates (Figure 1). For example, when we compare taxa pairwise, the substitution rates on lineages leading from the hominoid-OWM ancestor to hominoids are on average 2.68% (with a range of 2.63 to 2.74%) whereas rates on lineages leading to OWM are on average 3.66% (3.58 to 3.74%), 1.37-fold higher. These findings are qualitatively consistent with previous, smaller studies (31). Similarly, when considering the distance from the hominoid-NMW ancestor, substitution rates leading to NWM are on average 6.92% (6.89 to 6.94%), 1.64-fold higher than on the lineages leading to hominoids, which are on average 4.23% (4.18 to 4.29%). Substitution rates are also 1.62-fold higher in lineages leading to bushbaby (a prosimian) compared to hominoids (Figure 1). However, because of challenges in accurately reconstructing the ancestral state for

species that are closer to the outgroup, we believe this estimate to be less reliable and hence do not consider bushbaby in any further analysis.

Using bootstrap resampling of 1 Mb regions of the genome suggests tiny standard errors for the substitution rates (e.g., the standard error on lineages leading from the hominoid-OWM ancestor to hominoids is 0.01%), as expected from analyzing such large data sets. These standard errors are likely to be deceptively small, however, as the main source of uncertainty in our analysis is likely due to systematic effects of varying sequence quality, mapping and alignment artifacts among species. To evaluate the impact of these effects, we therefore repeat our analysis using a different sequence alignment of seven primates (the EPO dataset) and apply the same filters. When the species considered are matched between datasets, results are highly similar (Note S2), suggesting results are robust to data quality issues.

To evaluate how substitution patterns differ for mutations generated by distinct mechanisms, we distinguish between transitions at ancestrally CpG sites (referred to as CpG) outside of CpG islands (CGI), which are believed to occur mostly due to the spontaneous deamination of methylated cytosines, and transitions at ancestrally G or C sites outside of a CpG context (referred to as non-CpG G/C), which are thought to primarily occur due to replication errors. (Because CGI are often hypomethylated, we remove these regions from this analysis, and, unless specified otherwise, we refer to transitions at CpG sites outside of CGIs as "CpG transitions"). Mathematical modeling of different substitution mechanisms predicts that mutations that are non-replicative in origin and highly inefficiently repaired should depend on absolute time rather than on the number of cell divisions and hence should be more clock-like among species (25). In contrast, mutations that arise from replication errors or are non-replicative in origin but well repaired should depend on the generation time and other life history traits, and thus their substitution rates could vary considerably across primates (25, 59). Thus, a priori, we expect CpG transitions outside CGI to be more clock-like than other types of substitutions (assuming similar rates of deamination).

In order for our comparisons to not be confounded by biased gene conversion, we focus on the comparison of transitions at CpG sites with those occurring from non-CpG G/C sites. Because both types of mutations involve changes from G to A or C to T nucleotides, and both occur in regions with similar recombination rate profiles (Figure S2), they should, on average, be subject to similar strengths of biased gene conversion. Comparing hominoid and monkeys, the substitutions involving CpG transitions are on average 1.07-fold higher in lineages leading from the hominoid-OWM ancestor to OWM than they are in lineages leading to hominoids. Considering the hominoid-NWM ancestor, substitutions are 1.19-fold higher in lineages leading to NWM than to hominoids (Figure 2). In contrast, when considering transitions at non-CpG G/C sites, there are on average 1.39-fold more substitutions from the hominoid-OWM ancestor to OWM lineages than to hominoid ones and 1.71-fold more from the hominoid-NWM ancestor to NWM than to hominoid lineages (Figure 2). Thus, CpG transition rates are more similar across species, as observed in comparisons of smaller datasets of primates and mammals (7, 31). These results are robust to the choice of species of OWM and apes

10

used, e.g., using gorilla instead of chimpanzee or gibbon instead of orangutan yields similar differences (Figure S3, S4).

We then consider different substitution types in more detail, focusing on eight types: transitions and transversions occurring at either ancestrally A or T (referred to as A/T), ancestrally G or C (G/C), CpG and non-CpG G/C, again excluding CGI. As a measure of variation among species, we use the variance of the normalized root to leaf distances across all remaining nine species (see Methods), which is expected to be 0 if substitution rates are all identical. In general, transversions are more variable than transitions, with the largest variance (0.065) observed at A/T transversions (Figure 3a). In turn, the variance is lowest for CpG transitions outside of annotated CGI (0.005), as observed previously in comparisons of smaller datasets of 19 mammals (1.7 Mb) (7) and 9 primates (0.15 Mb) (60). Interestingly, transitions at CpG sites *inside* CGI have a greater variance in substitution rates, and behave similarly to G/C transitions (Figure 3a). The difference in behavior of CpGs inside and outside of CGI is again consistent with the notion that when the source of mutation is primarily non-replicative, mutations depend primarily on absolute time, whereas when they have other sources that are dependent on the numbers of cell divisions and are more variable among species. If this interpretation is correct, an interesting implication is that germline methylation levels and spontaneous deamination rates have remained very similar across primate species.

Patterns of substitutions may also vary across species due to effects of biased gene conversion, notably because of differences in effective population sizes (26). To examine this possibility, we compare the variance of the normalized root to leaf distances for substitutions that are subject to BGC (such as W-> S or S->W) and those that should not be affected by BGC (such as W->W and S->S). If the strength of BGC varies across primates, we expect larger variance across species at W->S and S->W substitutions. Instead, there is no significant difference in the estimates of variances across the two classes of substitutions (Figure 3b; p = 0.5, based on a permutation test). While this finding seems puzzling given the three to four-fold difference in effective population size of these species (28), it is consistent with results of Do et al. (2015), who found no significant difference in the extent of biased gene conversion across diverse groups of West African and non-African human populations that differ up to two-fold in their effective population sizes (61). If the strength of biased gene conversion at a site is typically very weak, both findings could reflect lack of power.

Given the importance of a steady molecular clock for dating events in human evolution, we next focus specifically on hominines. In these comparisons, subtle differences in sequence quality, coverage, or the extent of mapping artifacts can lead to misleading evidence for variation in substitution rates across species. To minimize these effects, we generated pairwise sequence alignments for high coverage (~30x) genomes of human and chimpanzee and human and gorilla. These pairs of genomes were mapped to the orangutan reference genome (which should be equidistant to all three species, assuming no differences in substitution rates among species), matching the alignment and variant calling pipeline for all three species (see Methods). In this way, and after removing

missing data, non-neutral sites and CGI, we obtain ~1.03 Gb for human-chimpanzee and ~1.02 Gb of human-gorilla whole genome sequence alignments.

Despite the differences in generation time and onset of puberty among extant chimpanzees and humans, rates of evolution on the two lineages are very similar: 0.621% and 0.633% respectively. This difference of 1.9% is however highly statistically significant under the assumption of no systematic errors ($p < 10^{-20}$). When we consider the substitution rates at different mutation types, there is somewhat more pronounced differences for some types of substitutions, in inconsistent directions. For example, when comparing chimpanzee to human branches for substitutions involving transversions from CpG sites, the difference is 0.91-fold, whereas it is 1.07-fold for transversions at A/T sites (Figure S9). Comparing human and gorilla lineages, differences in substitution rates are more pronounced: the gorilla branch (0.824%) is longer than that the human (0.773%) branch by an average of 6.6% ($p < 10^{-20}$). Again, different types of substitutions show distinct patterns, ranging between 0.96-fold at CpG transversions on the gorilla versus human branches to 1.10-fold for A/T transitions (Figure S10).

To check the reliability of these inferences, we also estimate the substitution rates using a second, newly developed maximum-likelihood based method (see Methods). While the absolute values of the substitution rates differ between the two methods, possibly due to methodological differences in calling ancestral states and assumptions about stationarity, the ratio of substitution rates between humans and chimpanzees (Figure S9, S11) or between humans and gorillas (Figure S10, S12) is almost identical. While these results match those obtained by Elango et al. (2005) based on 75 Mb of human-chimpanzee sequence, they are only about two-thirds of what was reported as the difference between humans and gorilla, based on 2 Mb of sequence data (32). Since our study is able to take advantage of a much larger dataset, accounts for differences in coverage and mapping among reference genomes and considers only putatively neutral sites, we surmise that the earlier estimate of the extent of substitution rate variation among human and gorilla was somewhat too high.

Importantly, these observations highlight that the mutation spectra, and not just the yearly mutation rate, are changing across primates. Notably, while the rate of substitutions involving CpG transitions is relatively stable across species, those of other substitution types vary substantially. As a result, the proportion of substitutions that consist in CpGs versus other mutation types differs among primates (Table S3). Since a transition to transversion ratio of 2.2, as found in humans, is often used as an indication of high quality genotypes (62), a practical implication is that this criterion cannot be reliably exported to other primates, or beyond. More fundamentally, our findings underscore that the mutation spectrum appears to have changed over the course of primate evolution. In this regard, it mirrors observations from even shorter time scales: for example, the recent report that transitions from 5'-TCC-3' → 5'-TTC-3' occurred at a proportionally higher rate in Europeans compared to Asians and Africans since these populations split (63).

Discussion

Evolutionary rates are faster in NWM compared to OWM and in turn; rates in OWM are faster than in humans and apes. These findings support the hominoid rate slowdown hypothesis (64, 65), indicating that since the split of hominoids and monkeys, per year mutation rates have decreased considerably. Moreover, the ordering of substitution rates is consistent with the generation time hypothesis, in that NWM have a substantially shorter generation time (g = ~6 years (20)) than OWM (g = ~11 years (66)), who in turn reproduce at younger ages than apes (g = ~25 years). Within hominines, gorillas (g = ~19 years (67)) have a faster yearly rate than humans (g = ~29 years) and chimpanzees (g = ~25 years (67)). To investigate if the association between generation time and substitution rates is significant after controlling for the underlying phylogeny, we perform independent contrast analysis (53). Specifically, we assume the underlying phylogeny based on CpG transition rates (effectively assuming these are strictly clock-like) and then estimate the correlation between generation times and non-CpG substitution rates, controlling for the shared phylogenetic history. While we have few species available for the analysis, the results are not significant (r = 0.17, p=0.7; see Methods), so the causal relationship remains to be established.

As an alternative approach is to ask whether differences in generation times and other life history traits can plausibly explain the variation in per year mutation rate. To this end, we consider a simple model for mutations that are replicative in origin (or non-replicative but well repaired (25)). Across primates, the average time between puberty and reproduction is positively correlated with age of onset of puberty in males (r = 0.74, p = 0.01, using Spearman's correlation test), and the rate of spermatogenesis (measured by estimating the seminiferous epithelium cycle length (SECL)) is positively correlated with generation time by (r = 0.90, p = 0.002) (Table S4). We therefore vary the generation time, age of onset of puberty and SECL for each lineage, relying on values estimated for extant humans, chimpanzees and OWM (Table S4). On that basis, we predict that yearly mutation rates in humans and chimpanzees should differ by ~30%, and hominoids and OWM should differ by ~66%. Thus, if anything, differences in life history traits predict even more variation in substitution rates than is observed.

Of course, these species may have had similar life histories throughout much of their evolutionary past, such that the use of life-history traits estimated in extant species exacerbates expected differences. For instance, fossil evidence suggests that the age of puberty on the human lineage may have only recently increased, and was lower in *Homo erectus* and Neanderthals (30, 68). If we change the age of onset of puberty in humans to 9 years, the difference in rates between humans and chimpanzees is only expected to be ~11%. One implication, then, of finding such similar substitution rates in humans and chimpanzees is that their life histories are likely to have been pretty similar for much of their evolutionary history.

That substitution rates should and do vary with life history traits highlights the challenges of using the molecular clock for dating evolutionary events, even within hominines. All the more so when there are substantial differences in the generation time between males and females (59). One way to overcome this difficulty is to explicitly model the changes in life-history traits over the course of primate evolution and study their effect on

substitution rates (59). Notably, Amster and Sella (2015) show that accounting for variation in generation time, age of onset of puberty and rate of spermatogenesis in apes helps to reconcile the split times estimated based on molecular and fossil evidence (59). This approach, however, requires knowledge of life-history traits in extant and ancestral populations, which is hard to obtain for many species.

An alternative may be to focus on mutation types that are much less sensitive to life-history traits, such as CpG transitions outside CGIs. Even for this mutation type, the variance in substitution rates across species is non-zero, possibly because a subset of these mutations occurs at unmethylated sites or through replication errors, or because repair is not completely inefficient. Nonetheless, CpG transitions accumulate in a quasi-clock like manner and appear to be least affected by life history differences across species. Moreover, in humans, they contribute almost a fifth of all *de novo* mutations (12).

With these considerations in mind, we re-estimate the divergence and split times of human, chimpanzees and gorillas using substitution rates estimated only at CpG transitions. Assuming the per year mutation rate for CpG transitions obtained in (12) (see Methods), we estimate that humans diverged from chimpanzees ~12.1 million years ago (Mya) and from gorillas ~15.1 Mya. Assuming further that the effective population size of the human-ape ancestor was five times the current population size (as estimated by (37, 56)), the human-chimpanzee split time is approximately 7.9 Mya and human-gorilla split time of 10.8 Mya. Reassuringly, these estimates are similar to those obtained by explicitly modeling the dependence of replicative mutations on life history traits in hominines (59). Moreover, these estimates are in broad agreement with evidence from the fossil record, which suggests a human-chimpanzee split time of 6-10 Mya and human-gorilla split time of 7-12 Mya (69-72). Thus, perhaps there is no real discrepancy between phylogenetic and pedigree based estimates of mutation rates, once the impact of life history traits on mutation rates is taken into account (59).

Acknowledgements

Figures

Figure 1: Phylogenetic trees for 10 primates. We estimate neutral substitution rates for 10 primates and an outgroup (mouse, shown in gray) from the Multiz dataset using Phylofit (see Methods for details of data set and filtering). Branch lengths reflect the expected number of neutral substitutions per site along each lineage.
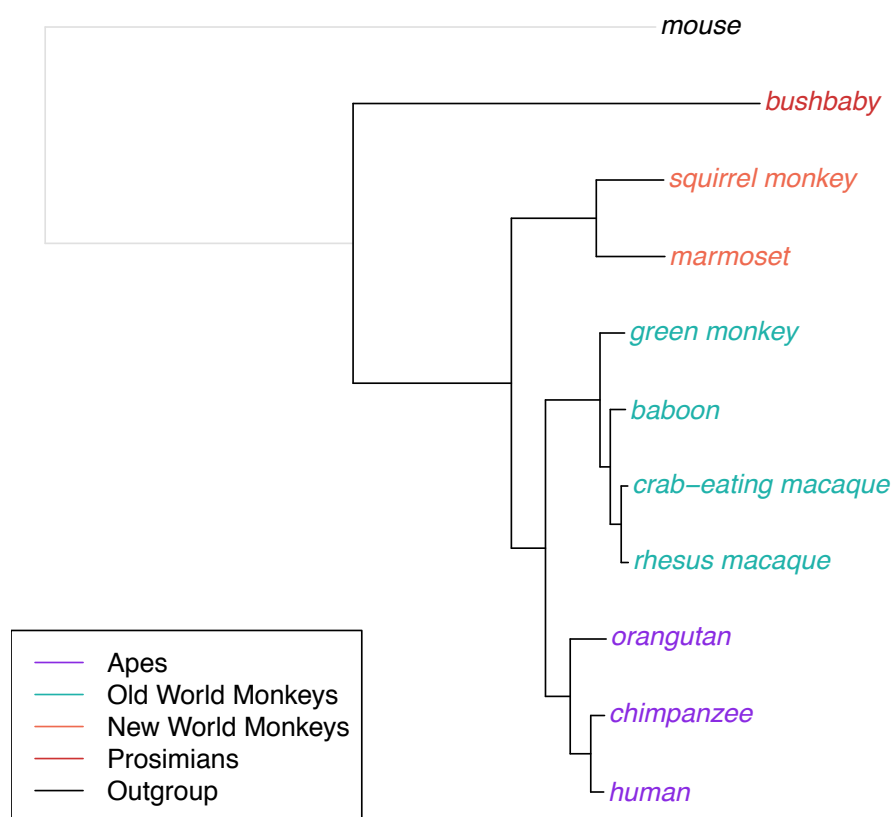
Figure 2: Comparison of substitution rates in hominoids and monkeys. For transitions from CpG and non-CpG G/C sites, we estimate the total branch length from either (a) the hominoid-OWM ancestor to each leaf, or (b) the hominoid-NWM ancestor to each leaf. The mean length for monkeys (OWM, NWM) and hominoids are shown, along with range within each group. Branches from root-hominoids are shown in purple, from root-OWM in green and from root-NWM in orange. The symbol * indicates the hominoid-monkey (either OWM or NWM) ancestor used as root.

(a) Hominoid–OWM

(b) Hominoid–NWM

16
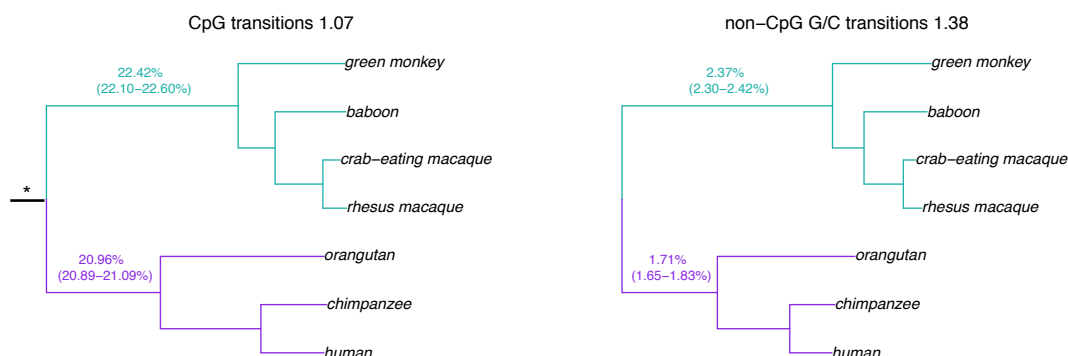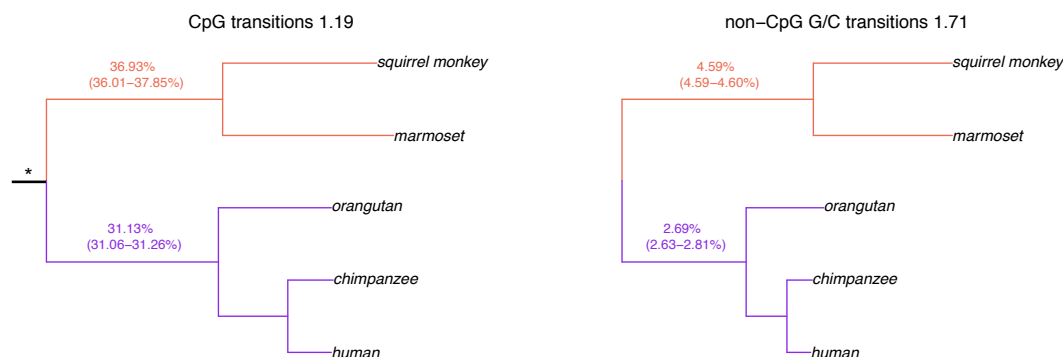
Figure 3: Variance among lineages for different substitution types. (a) For each ancestral state and each context shown on the x-axis, we estimate the total branch length from the root to each terminal leaf in the Multiz dataset as the inferred number of substitutions per site. We then compute the variance in the normalized root to leaf distance across nine primate species (human, chimpanzee, orangutan, rhesus macaque, crab-eating macaque, baboon, green monkey, squirrel monkey and marmoset). (b) For each substitution type (strong (S; G/C) and weak (W; A/T)) in different substitution contexts shown on the x-axis, we estimate the total branch length from the root to each terminal leaf in the Multiz dataset, and compute the variance in the root-leaf distance across the nine primates used in (a).

(a) variation in substitution rates, by mutation type and context
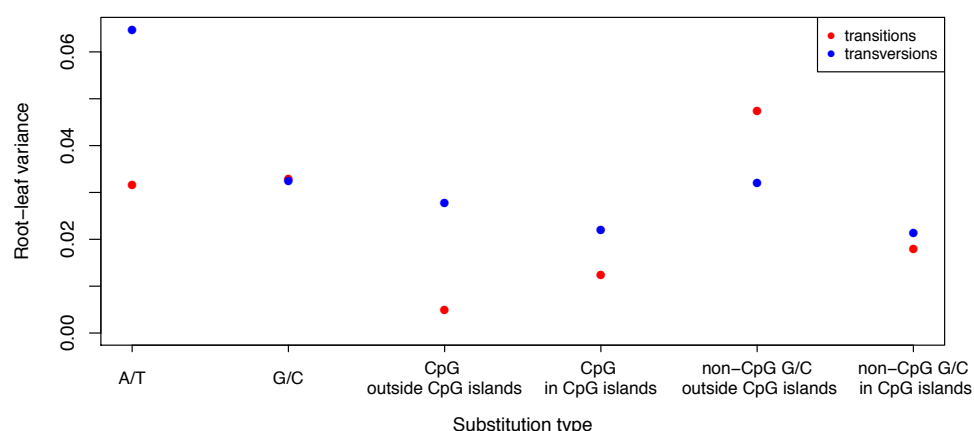
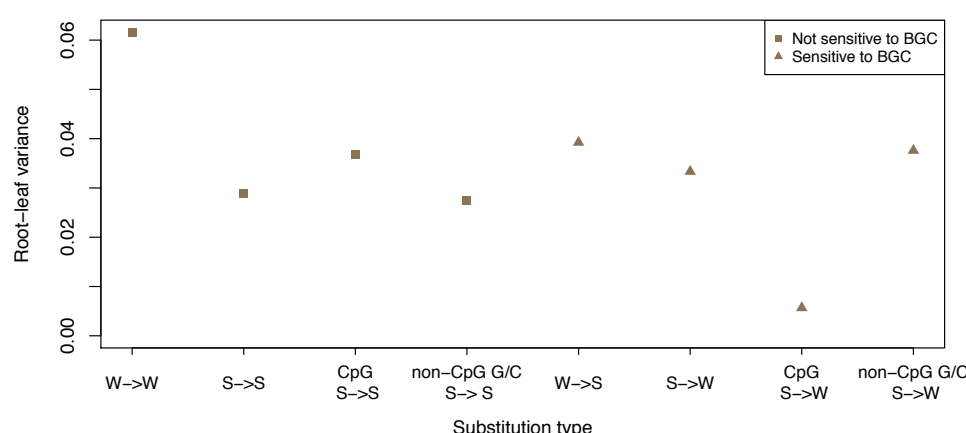**(b) variation in substitution rates, for types subject and not subject to BGC**

17

Figure 4. Revised estimates for divergence of hominines. We estimate the autosomal substitution rates for transitions at CpG sites by applying Phylofit to the high coverage pairwise alignment of human, chimpanzee and human and gorilla mapped to the orangutan reference genome. Divergence time is estimated from rates of *de novo* CpG transitions obtained in human pedigrees (see Methods for details).

(a)   **Human–chimpanzee divergence:12.1 Mya**   **(b)**   **Human–gorilla divergence: 15.1 Mya**
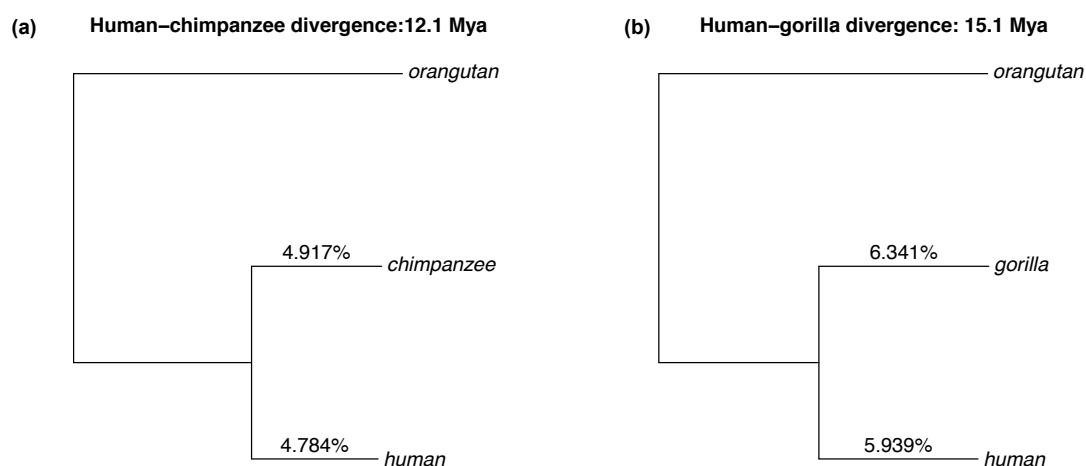
Supplementary material

Note S1: Generating high coverage data for human genome. For our high coverage analysis, we sequenced one individual of European ancestry. This individual provided informed consent for participation in the study. The project was approved by the Institutional Review Boards (IRBs) at The University of Chicago and Columbia University.

Genomic DNA was extracted from blood and one library was constructed using Illumina-recommended protocols. Briefly, 1μg of DNA was extracted and sheared into fragments using sonication. The resulting fragments were end repaired, adenosine overhangs were added and adaptors were ligated. Gel electrophoresis was performed to select libraries with insert sizes of approximately 350 bp in size, which were amplified using quantitative PCR. The resulting libraries were sequenced using Illumina HiSeq 2000 to generate paired-end reads. We generated ~49 Gb of sequencing data (~30x coverage). Mapping and alignment were done using samtools as described in the main text.

Sequence data will be available upon publication from dbGap from the following link: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000599.v2.p1

Note S2: Analysis of EPO dataset

To test the robustness of our inferences, we repeat 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 (51), we exclude gorilla from further analysis. After filtering putatively selected sites and removing missing data, we analyze approximately 745 Mb of whole genome sequence alignment. To allow for direct comparison with the Multiz dataset, we repeat 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 apply Phylofit to estimate the substitution rates across all species (Figure S5) and find that substitution rates on lineages leading from the hominoid-OWM ancestor to hominoids are on average 2.85% (range 2.81-2.88% across species) whereas rates on lineages leading to OWM are on average 3.59% (3.589-3.593%), 1.26-fold higher. These estimates are lower than results reported in the main text, as we are using a smaller subset of species. Similar estimates are obtained for the smaller subset of species with the Multiz dataset where the substitution rates are 1.28-fold faster in OWM compared to hominoids. We also repeat the main analyses shown in Figure 2 and 3 with the smaller subset of species in the EPO and Multiz dataset (Figure S6-S8).

19

Figure S1: Sperm methylation profiles of CpG sites. We plot the distribution of methylation levels at CpG sites inside and outside of annotated CGI. The methylation profiles in human sperm were taken from (47).
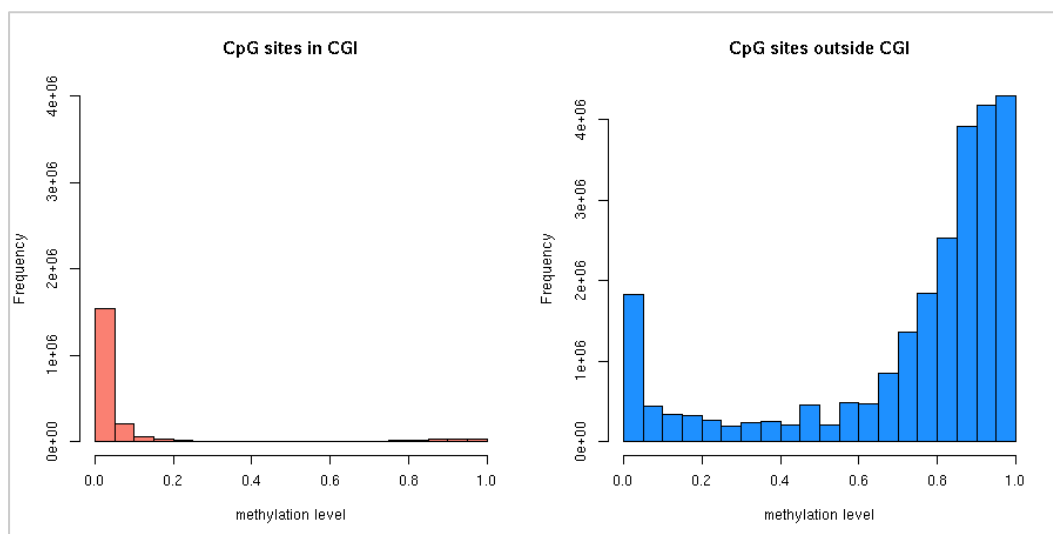

Figure S2: Distribution of CpG and non-CpG G/C sites across the human genome. We plot the proportion of CpG and non-CpG G/C sites in the human genome as a function of recombination rate. After filtering non-neutral sites and CGI (see Methods) in the Multiz dataset, we estimate the overall proportion of CpG and non-CpG G/C sites as 1.60% and 37.9% respectively. Crossover rates were obtained from the deCODE genetic map, UCSC genome browser track "deCODE Recombination maps: Sex avg" (73) originally calculated in 10 kb bins and standardized to have an average rate of 1.
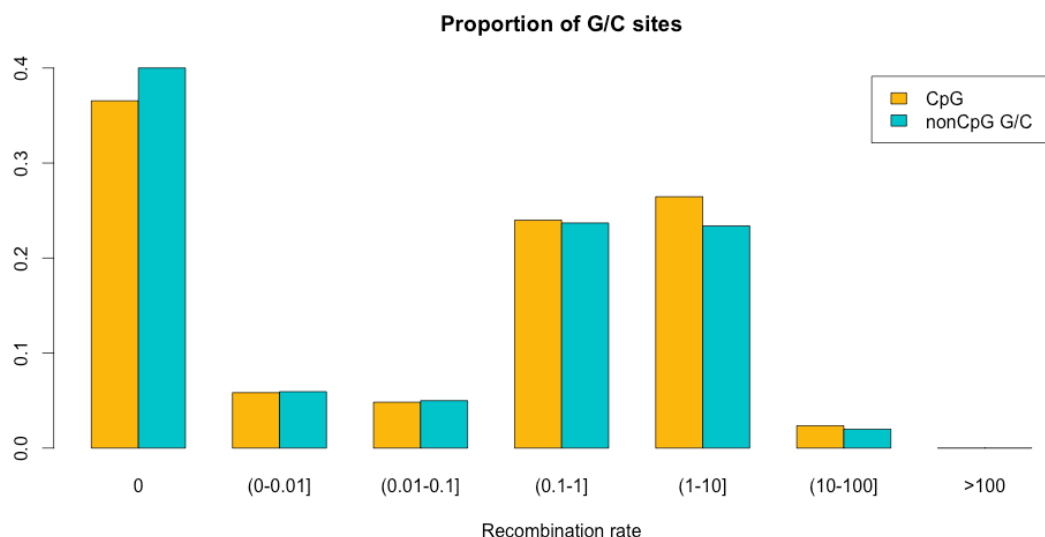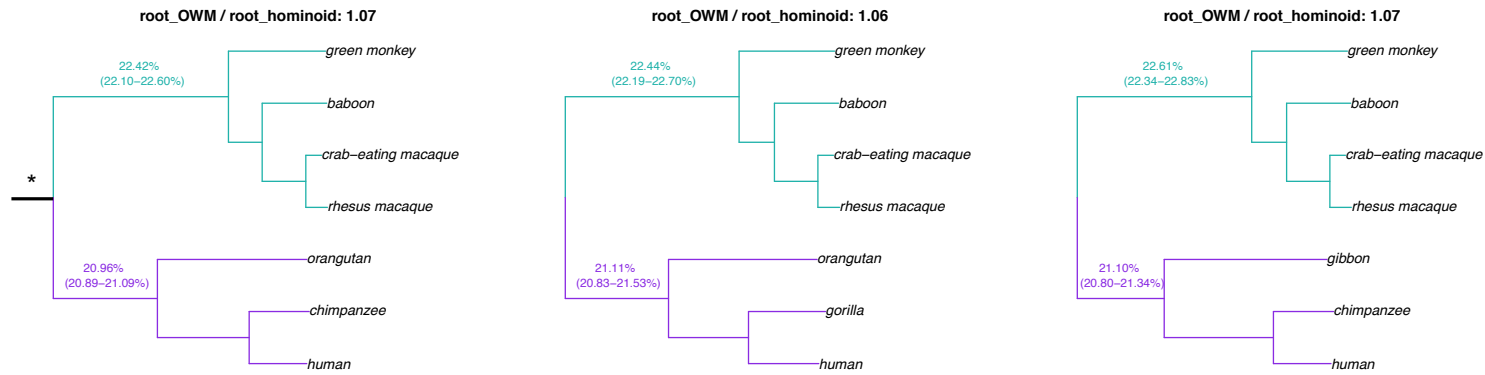
Figure S3: Comparison of substitution rates in hominoids and OWM using alternate topologies. Due to concerns about the possible effects of incomplete lineage sorting, we analyze gorilla and chimpanzees and gibbon and orangutans separately. Each sub-figure shows a different set of species and substitution type (CpG / non-CpG G/C sites). For each topology, we estimate the total branch length from the hominoid-OWM ancestor to each leaf. The mean lengths for OWMs and hominoids are shown, along with range within each group. Branches from root to hominoids are shown in purple and from root to OWMs are shown in green. The symbol * indicates the hominoid-OWM root from which branch lengths were estimated. The ratio of the average substitution rate leading to an OWM to the average rate leading to a hominoid is shown as the title for each sub-figure.

(a) Transitions at CpG sites

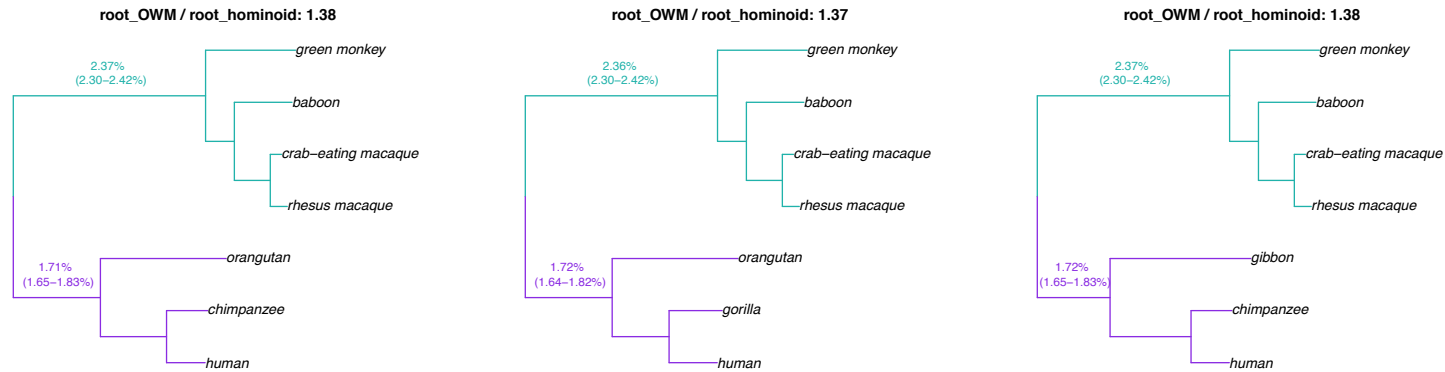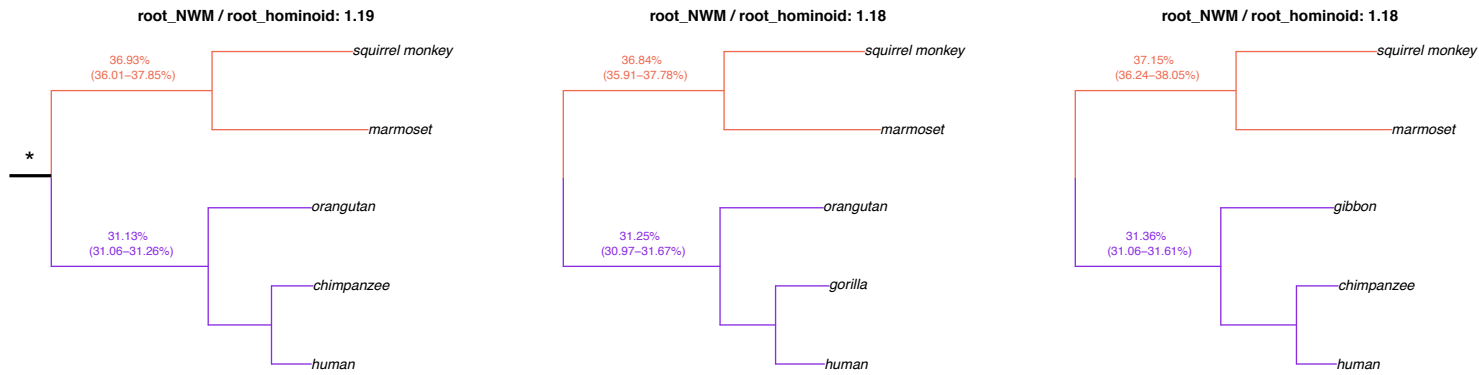(b) Transitions at non-CpG G/C sites

Figure S4: Comparison of substitution rates in hominoids and NWM using alternate topologies. Due to concerns about the possible effects of incomplete lineage sorting, we analyze gorilla and chimpanzees and gibbon and orangutans separately. Each sub-figure shows a different set of species and substitution type (CpG / non-CpG G/C sites). For each topology, we estimate the total branch length from the hominoid-NWM ancestor to each leaf. The mean lengths for NWMs and hominoids are shown, along with range within each group. Branches from root to hominoids are shown in purple and from root to NWMs are shown in green. The symbol * indicates the hominoid-NWM root from which branch lengths were estimated. The ratio of the average substitution rate leading to an NWM to the average rate leading to a hominoid is shown as the title for each sub-figure.

(a) Transitions at CpG sites
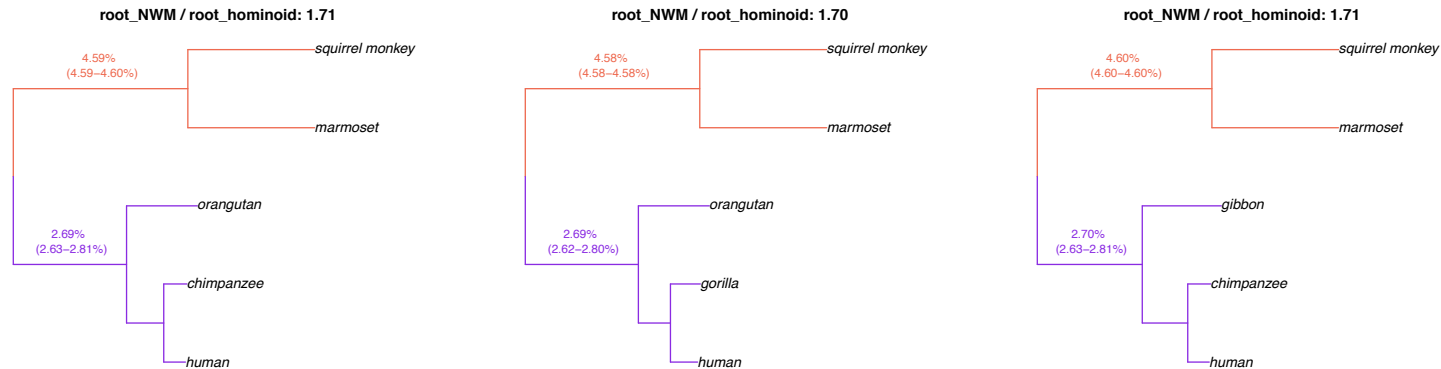
(b) Transitions at non–CpG G/C sites

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 Methods for details). Branch lengths reflect the expected number of neutral substitutions per site along each lineage. We excluded gorilla due to concerns of possible effects of incomplete lineage sorting on estimates of substitution rates.
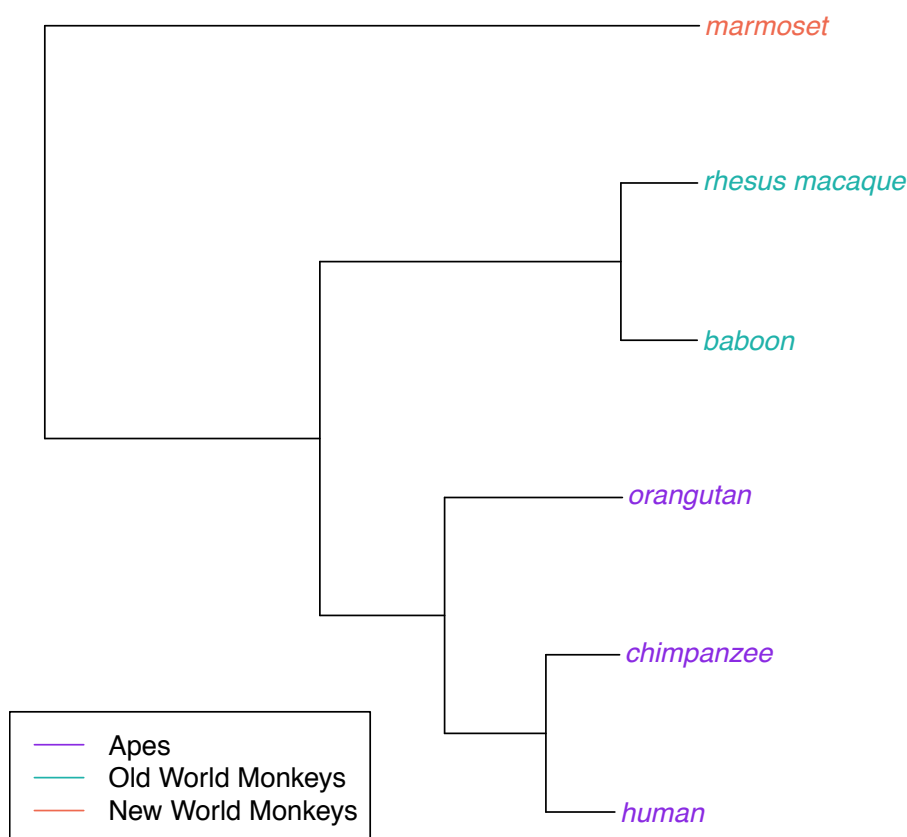
Figure S6: Comparison of substitution rates in hominoids and OWM using different datasets. For each dataset (Multiz or EPO), we estimate the total branch length from the hominoid-OWM ancestor to each leaf. The mean lengths for OWM and hominoids are shown, along with the range within each group. Branches from root to hominoids are shown in purple and from root to OWM are shown in green. The ratio of the average substitution rate on branches leading to OWMs to the average rate on branches leading to hominoids is shown as the title for each sub-figure along with the substitution context.
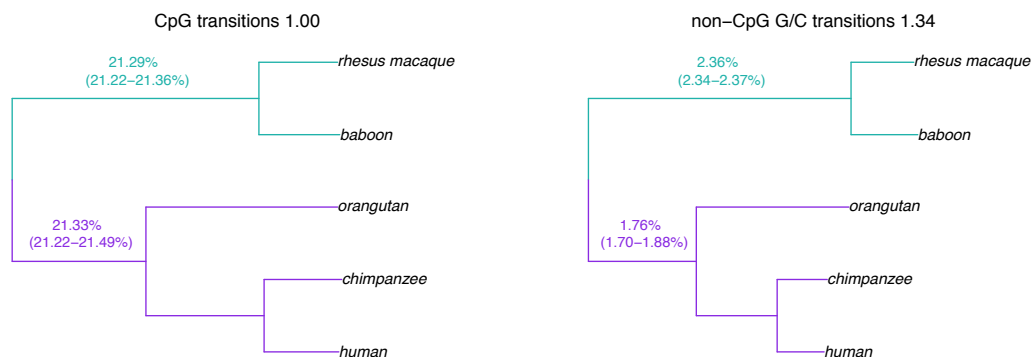

(a) Multiz
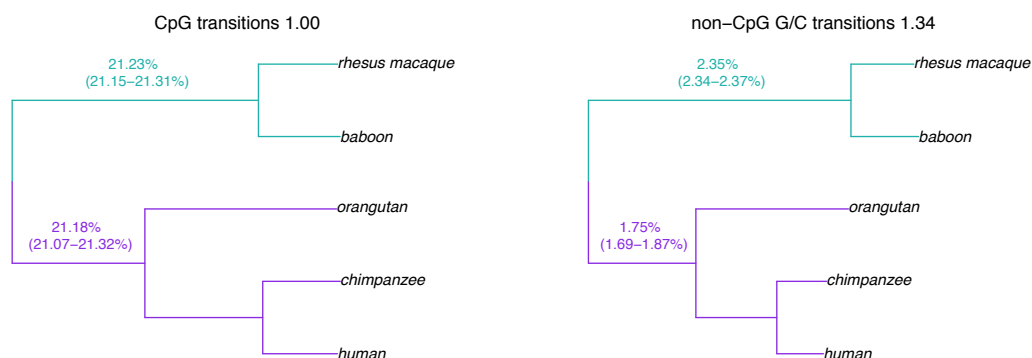
(b) EPO

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 estimate the total branch length from the root to each terminal leaf in the Multiz and EPO dataset as the inferred number of substitutions per site. We then compute 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 types of species (hominoids and OWM) available in the EPO dataset.

(a) Multiz
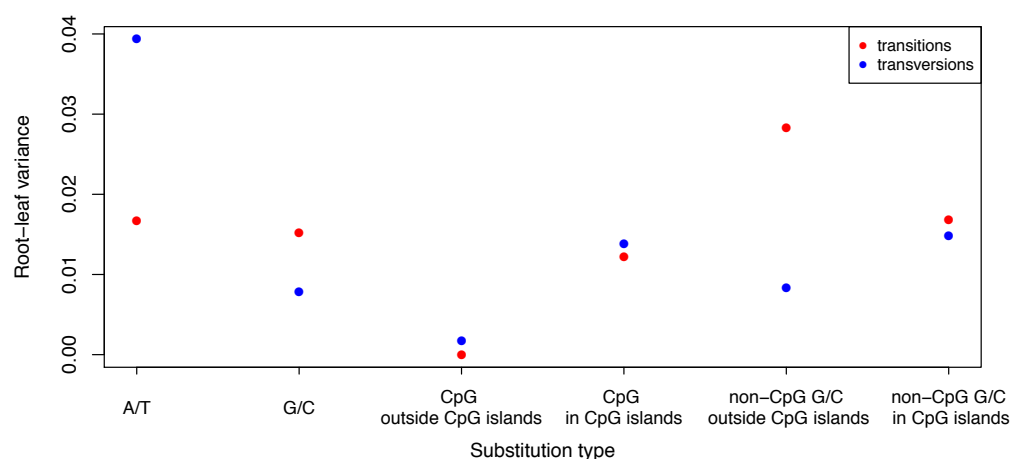
(b) EPO
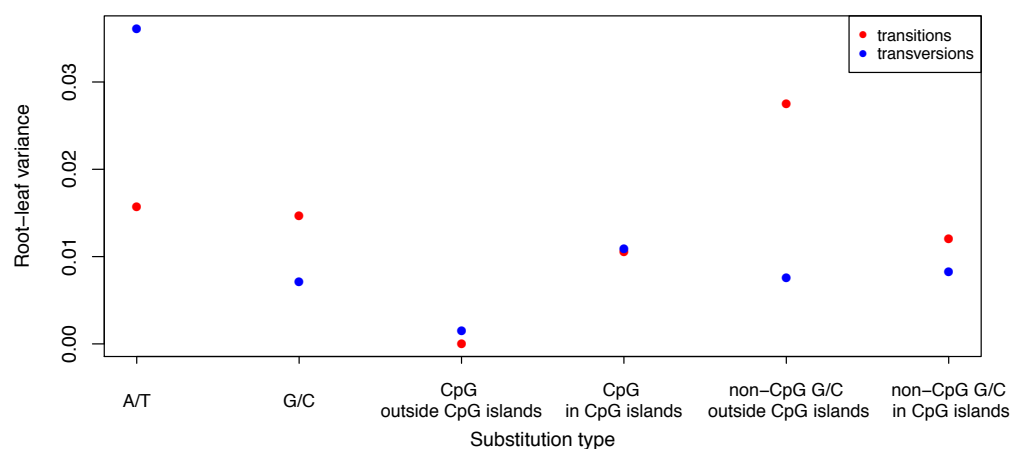
Figure S8: Effect of biased gene conversion across lineages estimated for different datasets. For each substitution type (strong (S; G/C) and weak (W; A/T)) and each ancestral context shown on the x-axis, we estimate the total branch length from the root to each terminal leaf in the Multiz and EPO dataset as the inferred number of substitutions per site. We then compute the variance in the normalized root to leaf distance across five primates in EPO (human, chimpanzee, orangutan, rhesus macaque and baboon). This figure differs from Figure 2b, as it uses fewer species in the Multiz dataset to match the types of species (hominoids and OWM) available in the EPO dataset.
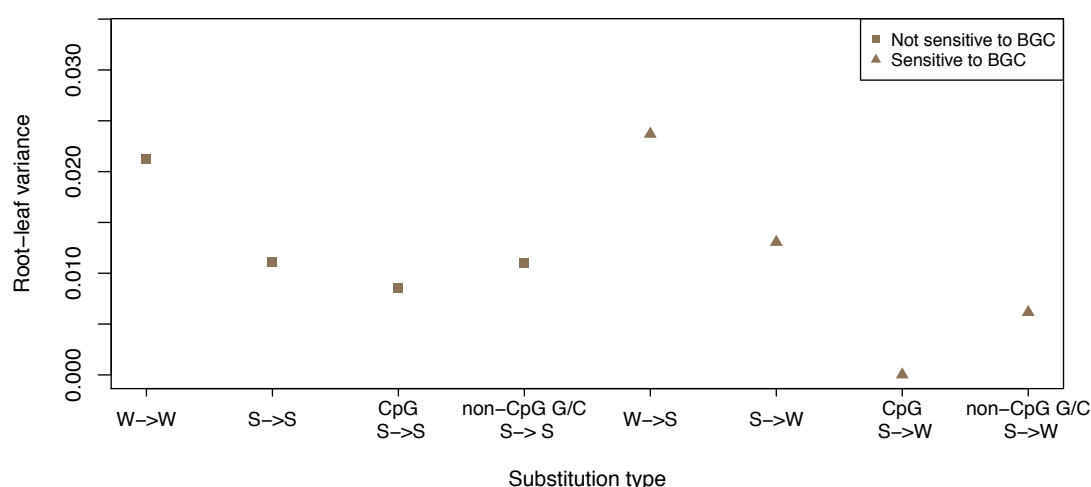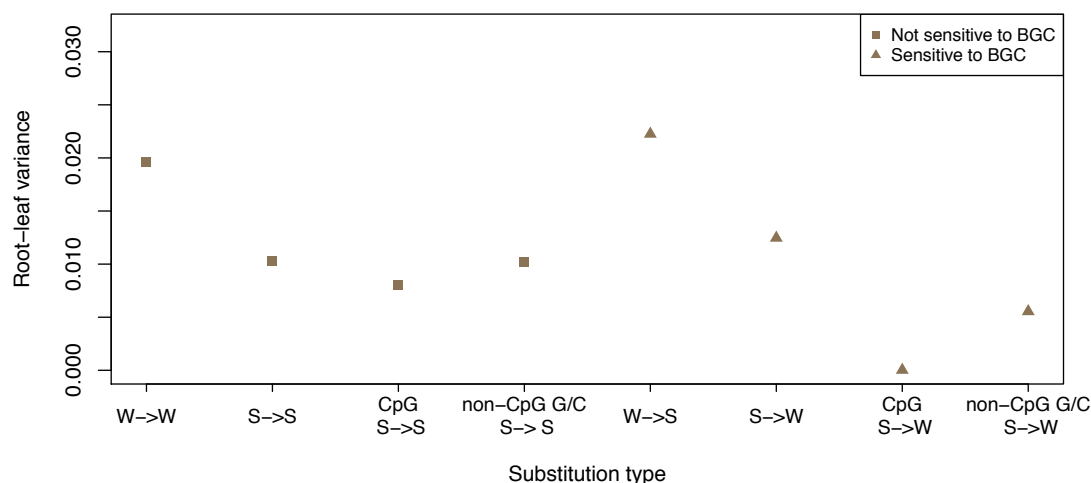
(a) Multiz

(b) EPO

26

Figure S9: Comparison of substitution rates in humans and chimpanzees using Phylofit. For each substitution type, we estimate the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio of substitution rate in chimpanzee to the substitution rate in human is shown as the title of each subfigure.

(a) Transitions

ancestral A/T: 1.01	ancestral C/G: 1.02	ancestral CpG: 1.03	ancestral non−CpG G/C: 1.02
0.911% orangutan	1.562% orangutan	12.431% orangutan	1.098% orangutan
0.343% chimpanzee	0.536% chimpanzee	4.917% chimpanzee	0.353% chimpanzee
0.496%	0.747%	6.728%	0.494%
0.341% human	0.524% human	4.784% human	0.346% human

(b) Transversions

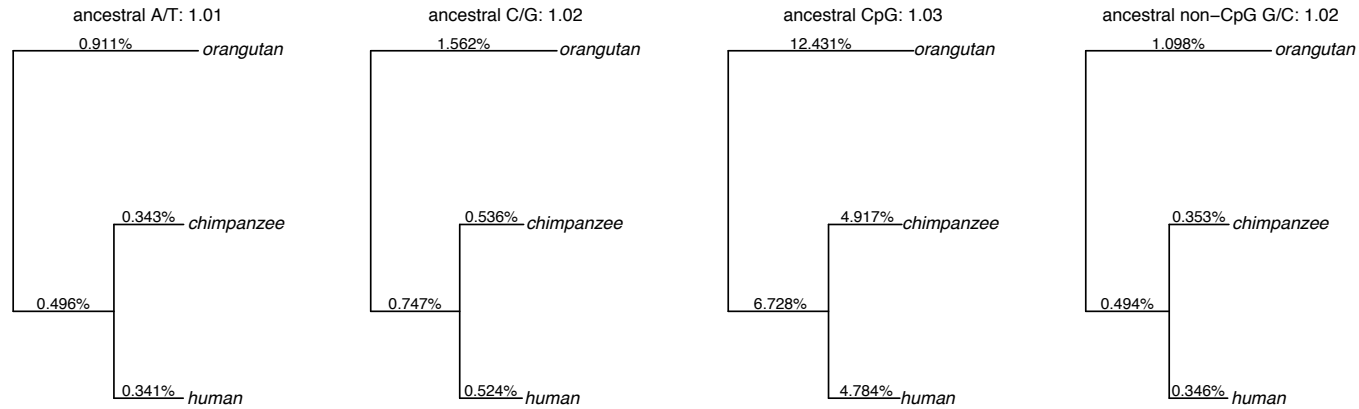| ancestral A/T: 1.07	ancestral C/G: 1.00	ancestral CpG: 0.91	ancestral non−CpG G/C: 1.01
0.408% orangutan	0.707% orangutan	1.479% orangutan	0.672% orangutan
0.166% chimpanzee	0.256% chimpanzee	0.529% chimpanzee	0.244% chimpanzee
0.216%	0.347%	0.774%	0.329%
0.155% human	0.256% human	0.582% human	0.242% human

27

Figure S10: Comparison of substitution rates in humans and gorillas using Phylofit. For each substitution type, we estimate the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of substitution rate in gorilla to the substitution rate in human is shown as the title of each subfigure.

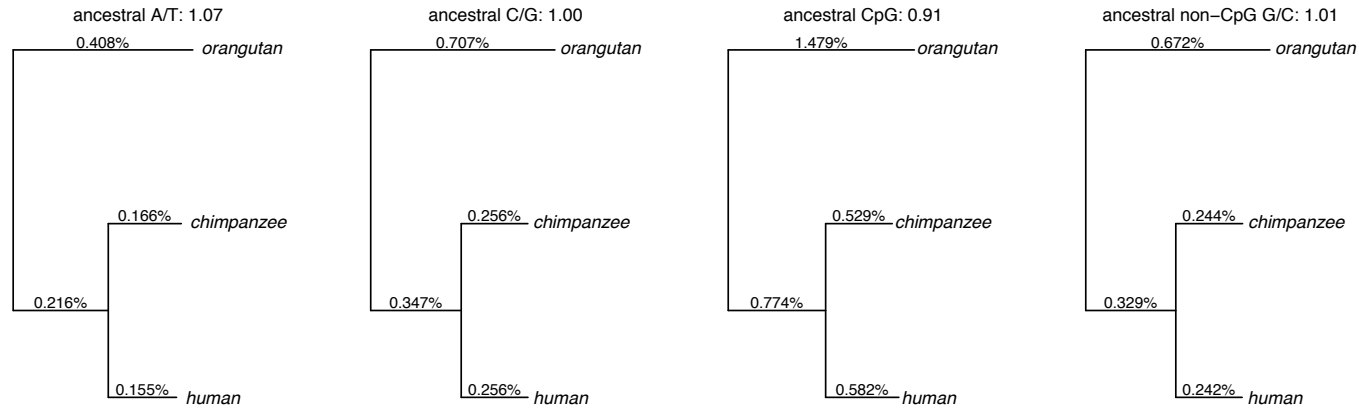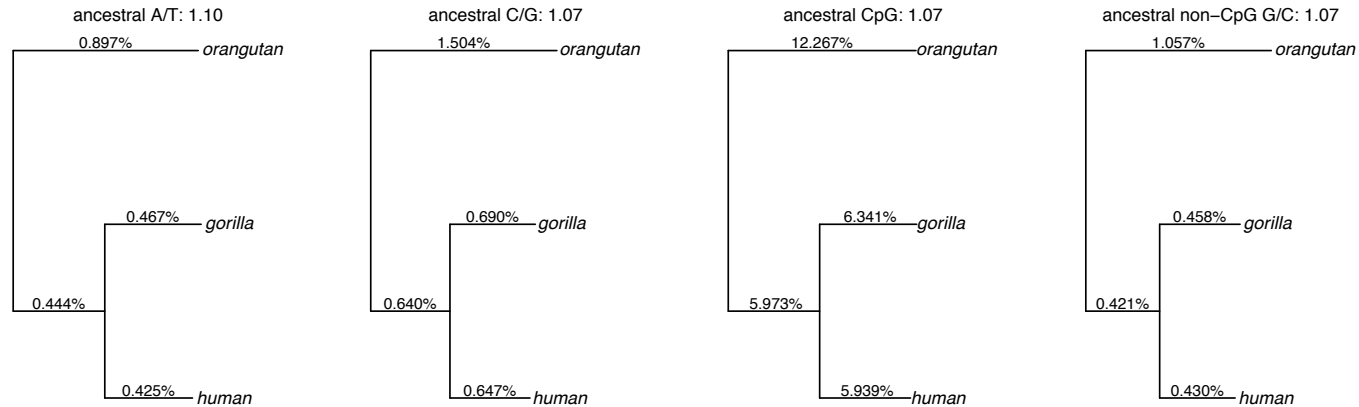**(a) Transitions**

(b) Transversions

Figure S11: Comparison of substitution rates in humans and chimpanzees using the maximum likelihood approach. For each substitution type, we estimate the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio of substitution rate in chimpanzee to the substitution rate in human is shown as the title of each subfigure. We note the maximum likelihood approach does not estimate the rates for ancestral C/G sites (that include CpG) and hence we do not show results for this context.
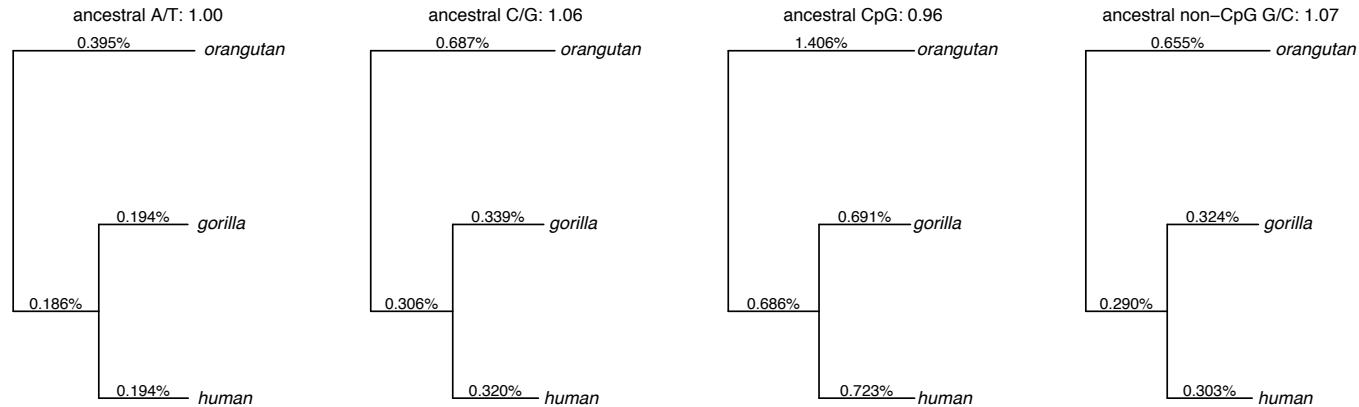
Figure S12: Comparison of substitution rates in humans and gorillas using the maximum likelihood approach. For each substitution type, we estimate the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of substitution rate in gorilla to the substitution rate in human is shown as the title of each subfigure. We note the maximum likelihood ap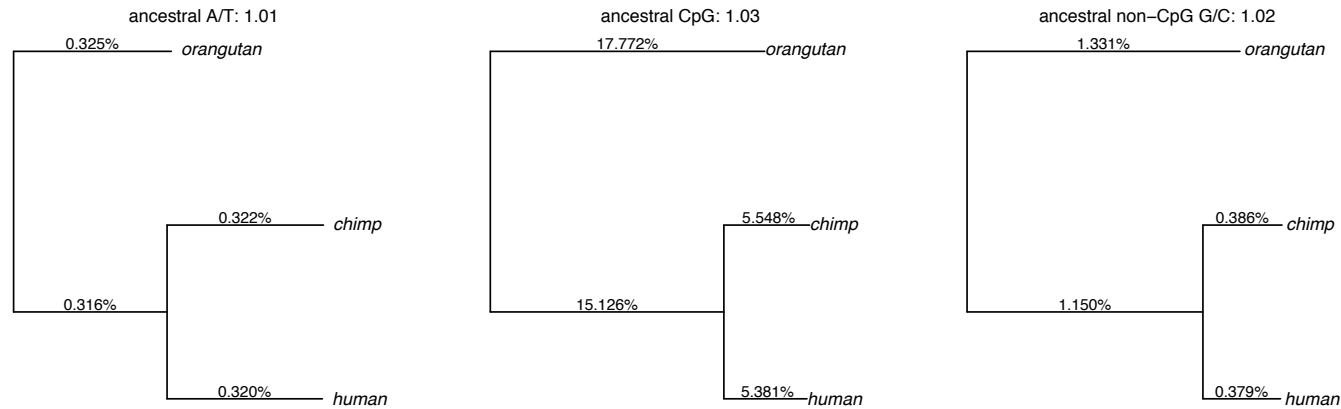proach does not estimate the rates for ancestral C/G sites (that include CpG) and hence we do not show results for this context.
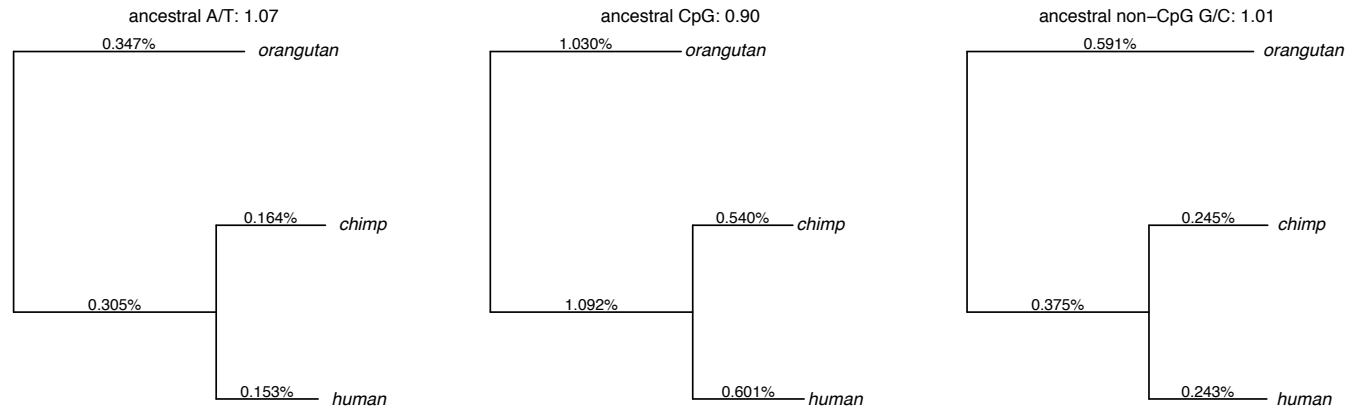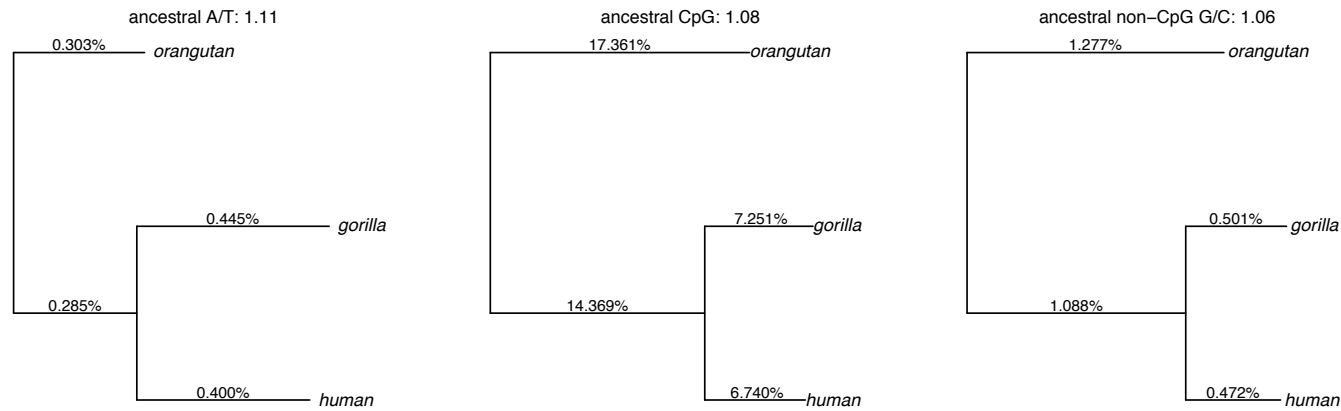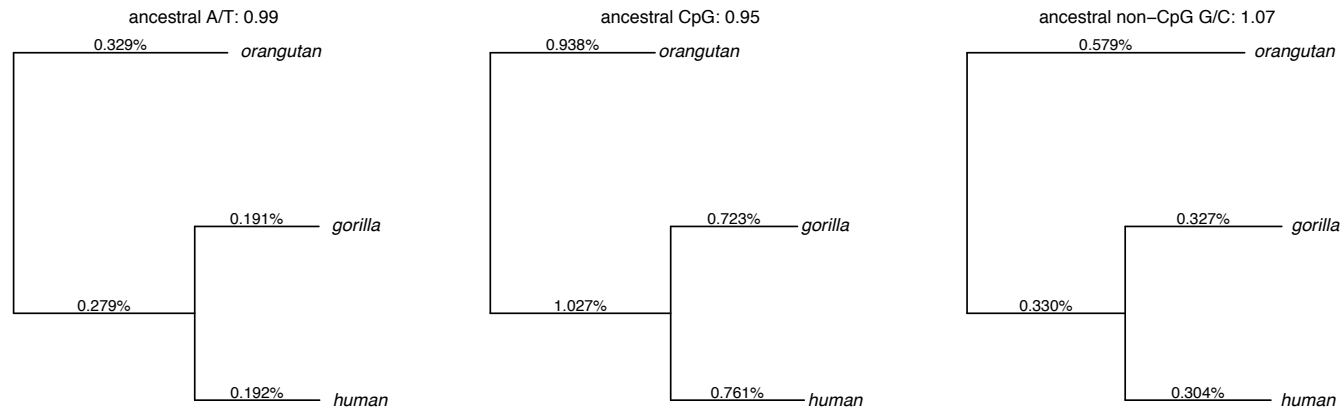
Table S1. Online source of annotation for transposable elements, coding exons, CGI, and conserved sites.

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: Life history traits in primates

Species	Common name	SECL (in days)[a]	Generation time (in years)	Onset of puberty (in years)[b]
Homo sapiens	Human	16 (74)	29 (21)	13.5 (75)
Pan troglodytes	Chimp	14 (76)	25 (67)	8.5 (77)
Gorilla gorilla	Gorilla	n/a	19 (67)	7 (78)
Pongo abelii	Orangutan	n/a	27* (79)	6.5[+] (80)
Macaca fascicularis	Crab-eating macaque	10.2 (81)	11[+] (82)	3.5 (83)
Macaca mulatta	Rhesus macaque	10.5 (84)	12 (82)	3.5 (83)
Papio anubis	Baboon	11 (85)	11 (66)	5.4[+] (86)
Cercopithecus aethiops	Green Monkey	10.2 (87)	11[+] (88)	5 (83)
Saimiri sciureus	Squirrel Monkey	10.2 (87)	9* (89)	3 (90)
Callithrix jacchus	Marmoset	10 (91)	6 (20)	0.9 (92)

Note: n/a = not available. * only female generation was available. [+] inferred from a closely related species.
[a] source: (22)
[b] main source: (90)

Table S3. Proportion of substitution in each context across lineages.

Species	A/T transitions	A/T transversions	G/C transitions	G/C transversions	CpG transitions	CpG transversions	non-CpG G/C transitions	non-CpG G/C transversions
hg19	0.344	0.152	0.340	0.164	0.118	0.014	0.221	0.150
panTro4	0.333	0.163	0.339	0.164	0.116	0.012	0.222	0.151
hg19-panTro4	0.334	0.150	0.357	0.158	0.119	0.013	0.237	0.145
ponAbe2	0.319	0.145	0.373	0.162	0.112	0.012	0.259	0.150
hg19-ponAbe2	0.303	0.160	0.380	0.157	0.118	0.014	0.261	0.142
rheMac3	0.289	0.170	0.377	0.164	0.124	0.013	0.252	0.150
macFas5	0.302	0.162	0.387	0.149	0.136	0.014	0.251	0.135
rheMac3-macFas5	0.315	0.158	0.383	0.144	0.123	0.014	0.259	0.130
papHam1	0.301	0.156	0.399	0.144	0.121	0.013	0.277	0.131
rheMac3-papHam1	0.357	0.159	0.355	0.129	0.111	0.012	0.244	0.117
chlSab1	0.306	0.167	0.382	0.145	0.112	0.013	0.268	0.131
rheMac3-chlSab1	0.320	0.176	0.357	0.148	0.091	0.012	0.264	0.135
hg19-rheMac3	0.311	0.178	0.358	0.152	0.106	0.016	0.251	0.136
calJac3	0.321	0.185	0.345	0.149	0.096	0.014	0.246	0.134
saiBol1	0.332	0.182	0.340	0.146	0.087	0.013	0.250	0.132
calJac3-saiBol1	0.291	0.193	0.365	0.152	0.082	0.014	0.280	0.137

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. 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.

2. Kimura M (1984) *The neutral theory of molecular evolution* (Cambridge University Press).

3. Zuckerkandl E & Pauling L (1965) Evolutionary divergence and convergence in proteins. *Evolving genes and proteins* 97:97-166.

4. Morgan GJ (1998) Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *Journal of the History of Biology* 31(2):155-178.

5. Kumar S (2005) Molecular clocks: four decades of evolution. *Nature Reviews Genetics* 6(8):654-662.

6. Sayres MAW, Venditti C, Pagel M, & Makova KD (2011) Do Variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution; international journal of organic evolution* 65(10):2800-2815.

7. 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.

8. Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231(4744):1393-1398.

9. Wu C-I & Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences* 82(6):1741-1745.

10. Nachman MW & Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156(1):297-304.

11. Sequencing TC & Consortium A (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437(7055):69-87.

12. 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.

13. Campbell CD, *et al.* (2012) Estimating the human mutation rate using autozygosity in a founder population. *Nature genetics* 44(11):1277-1281.

14. Scally A & Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics* 13(10):745-753.

15. Martin AP & Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences* 90(9):4087-4091.

16. Fontanillas E, Welch JJ, Thomas JA, & Bromham L (2007) The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC evolutionary biology* 7(1):95.

17. Wong A (2014) Covariance between testes size and substitution rates in primates. *Molecular biology and evolution*:msu091.

18.   Drost JB & Lee WR (1995) Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environmental and molecular mutagenesis* 25(S2):48-64.

19.   Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics* 1(1):40-47.

20.   Gage TB (1998) The comparative demography of primates: with some comments on the evolution of life histories. *Annual Review of Anthropology*:197-221.

21.   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.

22.   Ramm SA & Stockley P (2009) Sperm competition and sperm length influence the rate of mammalian spermatogenesis. *Biology letters*:rsbl20090635.

23.   Coulondre C, Miller JH, Farabaugh PJ, & Gilbert W (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* 274(5673):775-780.

24.   Duncan BK & Miller JH (1980) Mutagenic deamination of cytosine residues in DNA.

25.   Gao Z, Wyman MJ, Sella G, & Przeworski M (2015) Interpreting the dependence of mutation rates on age and time. *arXiv preprint arXiv:1507.06890*.

26.   Duret L & Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics* 10:285-311.

27.   Capra JA, Hubisz MJ, Kostka D, Pollard KS, & Siepel A (2013) A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet* 9(8):e1003684.

28.   Leffler EM*, et al.* (2012) Revisiting an old riddle: what determines genetic diversity levels within species. *PLoS biology* 10(9):e1001388.

29.   Stevison LS*, et al.* (2015) The Time-Scale of Recombination Rate Evolution in Great Apes. *bioRxiv*:013755.

30.   Auton A*, et al.* (2012) A fine-scale chimpanzee genetic map from population sequencing. *Science* 336(6078):193-198.

31.   Kim S-H, Elango N, Warden C, Vigoda E, & Soojin VY (2006) Heterogeneous genomic molecular clocks in primates. *PLoS genetics* 2(10):e163.

32.   Elango N, Thomas JW, & Soojin VY (2006) Variable molecular clocks in hominoids. *Proceedings of the National Academy of Sciences of the United States of America* 103(5):1370-1375.

33.   Patterson N, Richter DJ, Gnerre S, Lander ES, & Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441(7097):1103-1108.

34.   Karolchik D*, et al.* (2014) The UCSC genome browser database: 2014 update. *Nucleic acids research* 42(D1):D764-D770.

35.   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.

36.   Venn O*, et al.* (2014) Strong male bias drives germline mutation in chimpanzees. *Science* 344(6189):1272-1275.

37. Prado-Martinez J, *et al.* (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471-475.

38. Earl D, *et al.* (2014) Alignathon: a competitive assessment of whole-genome alignment methods. *Genome research* 24(12):2077-2089.

39. Locke DP, *et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469(7331):529-533.

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

41. Li H, *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

42. Siepel A, *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* 15(8):1034-1050.

43. Murphy WJ, *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294(5550):2348-2351.

44. 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.

45. Smit A, Hubley R, & Green P (2004) RepeatMasker Open-3.0. 2004. *Seattle (WA): Institute for Systems Biology*.

46. 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.

47. Molaro A, *et al.* (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146(6):1029-1041.

48. Gardiner-Garden M & Frommer M (1987) CpG islands in vertebrate genomes. *Journal of molecular biology* 196(2):261-282.

49. Siepel A & Haussler D (2004) Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Molecular Biology and Evolution* 21(3):468-488.

50. Duret L & Arndt PF (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* 4(5):e1000071.

51. Mailund T, Munch K, & Schierup MH (2014) Lineage sorting in apes. *Annual review of genetics* 48:519-535.

52. Jombart T & Dray S (2013) adephylo: exploratory analyses for the phylogenetic comparative method.

53. Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*:1-15.

54. Paradis E (2011) *Analysis of Phylogenetics and Evolution with R* (Springer Science & Business Media).

55. Gao J-J, *et al.* (2011) Highly variable recessive lethal or nearly lethal mutation rates during germ-line development of male Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 108(38):15914-15919.

56. Wall JD (2003) Estimating ancestral population sizes and divergence times. *Genetics* 163(1):395-404.

57. Meyer M, *et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-226.

58. Prüfer K, *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-49.

59. Amster G & Sella G (2015) Life history effects on the molecular clock of autosomes and sex chromosomes. *bioRxiv*:024281.

60. Lee H-J, Rodrigue N, & Thorne JL (2015) Relaxing the Molecular Clock to Different Degrees for Different Substitution Types. *Molecular biology and evolution*:msv099.

61. Do R, *et al.* (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature genetics* 47(2):126-131.

62. Consortium GP (2010) A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061-1073.

63. Harris K (2015) Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences* 112(11):3439-3444.

64. Goodman M (1963) Serological analysis of the systematics of recent hominoids. *Human biology*:377-436.

65. Soojin VY (2013) Morris Goodman's hominoid rate slowdown: the importance of being neutral. *Molecular phylogenetics and evolution* 66(2):569-574.

66. 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.

67. 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.

68. Gluckman P & Hanson M (2006) Changing times: the evolution of puberty. *Molecular and cellular endocrinology* 254:26-31.

69. Wilkinson RD, *et al.* (2010) Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic biology*:syq054.

70. Begun DR (2015) Fossil Record of Miocene Hominoids. *Handbook of Paleoanthropology*:1261-1332.

71. Brunet M, *et al.* (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418(6894):145-151.

72. Suwa G, Kono RT, Katoh S, Asfaw B, & Beyene Y (2007) A new species of great ape from the late Miocene epoch in Ethiopia. *Nature* 448(7156):921-924.

73. Kong A, *et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467(7319):1099-1103.

74. Heller C & Clermont Y (1963) Kinetics of the germinal epithelium in man. *Recent progress in hormone research* 20:545-575.

75. Marshall WA & Tanner JM (1970) Variations in the pattern of pubertal changes in boys. *Archives of disease in childhood* 45(239):13-23.

76. 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.

77. 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.

78. 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.

79. 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).

80. Dixson A, Knight J, Moore H, & Carman M (1982) Observations on sexual development in male Orang‐utans. *International Zoo Yearbook* 22(1):222-227.

81. 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.

82. 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).

83. Bercovitch FB (2000) Behavioral ecology and socioendocrinology of reproductive maturation in cercopithecine monkeys. *Old world monkeys*:298-320.

84. 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.

85. 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.

86. Onyango PO, Gesquiere LR, Altmann J, & Alberts SC (2013) Puberty and dispersal in a wild primate population. *Hormones and behavior* 64(2):240-249.

87. Barr A (1973) Timing of spermatogenesis in four nonhuman primate species. *Fertility and sterility* 24(5):381-389.

88. 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.

89. Abee CR, Mansfield K, Tardif SD, & Morris T (2012) *Nonhuman Primates in Biomedical Research: biology and management* (Academic Press).

90. Dixson AF (2009) *Sexual selection and the origins of human mating systems* (Oxford University Press).

91. 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.

92. 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.