# Variation in the molecular clock of primates

Priya Moorjani[a,b,1,2], Carlos Eduardo G. Amorim[a,1], Peter F. Arndt[c], and Molly Przeworski[a,d,2]

[a]Department of Biological Sciences, Columbia University, New York, NY 10027; [b]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142; [c]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; and [d]Department of Systems Biology, Columbia University, New York, NY 10027

Edited by David C. Page, Whitehead Institute, Cambridge, MA, and approved July 19, 2016 (received for review January 8, 2016)

**Events in primate evolution are often dated by assuming 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 it should also be found among primates. Motivated by these considerations, we analyze whole genomes from 10 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 up to 64% 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 clocklike behavior than do other types, presumably because of their nonreplicative 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 the human and chimpanzee divergence time is 12.1 million years, and the human and gorilla divergence time is 15.1 million years.**

molecular clock | mutation rate | primate evolution | CpG transition rate | human–ape divergence time

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 with 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 left Africa) 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).

However, 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 (5–7). In particular, there is the well-known hypothesis of a "generation time effect" on substitution rates, based on the observation that species with shorter generation time (i.e., mean age of reproduction) have higher mutation rates (8). For instance, mice have a generation time on the order of months (∼10–12 mo) compared with ∼29 y in humans (9), and a two- to threefold higher substitution rate per year (8). More generally, a survey of 32 mammalian species found reproductive span to be the strongest predictor of substitution rate variation (5).

A generation time effect has also been suggested in humans, motivated by the observation that the yearly mutation rate estimated by sequencing human and chimpanzee pedigrees [$\sim 0.4 \times 10^{-9}$ per base pair per year (10, 11)] is approximately twofold lower than the mutation rate inferred from the number of substitutions observed between primates (1). Substitution-derived estimates of mutation rates are highly dependent on dating evolutionary lineages from the fossil record, and so are subject to considerable uncertainty. Nonetheless, one way to reconcile pedigree and substitution-derived estimates of the mutation rate would be to postulate that the generation time has increased toward the present, and led to a decrease in the yearly mutation rate (12).

Whether the association between generation time and substitution rates is causal remains unclear, however; correlated traits such as metabolic rate (13), body size (14), and sperm competition (15) 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 (13). Body mass has been shown to be negatively correlated to substitution rates, such that smaller animals tend to have higher substitution rates (13). Sexual selection on mating systems may also affect substitution rates, as 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 (15).

That said, an effect of life history traits such as generation time on the yearly mutation rate is expected from first principles, given our understanding of oogenesis and spermatogenesis (16, 17). In mammals, oogonial divisions are completed by the birth of the future mother, whereas the spermatogonial stem cells continue to

## Significance

Much of our understanding of the chronology of human evolution relies on a fixed "molecular clock"; that is, a constant rate of substitutions per unit time. To evaluate the validity of this assumption, we analyze whole-genome sequences from 10 primate species. We find that there is substantial variation in the molecular clock between apes and monkeys and that rates even differ within hominines. Importantly, not all mutation types behave similarly; notably, transitions at CpG sites exhibit a more clocklike behavior than other substitutions, presumably because of their nonreplicative origin. Thus, the mutation spectra, and not just the overall substitution rates, are changing across primates. This finding suggests that events in primate evolution are most reliably dated using CpG transitions.

Author contributions: P.M., C.E.G.A., and M.P. designed research; P.M., C.E.G.A., and M.P. performed research; P.M. and P.F.A. contributed new reagents/analytic tools; P.M., C.E.G.A., and M.P. analyzed data; and P.M., C.E.G.A., and M.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]P.M. and C.E.G.A. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pm2730@columbia.edu or mp3284@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600374113/-/DCSupplemental.

divide postpuberty (16). 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, 17). These considerations indicate that changes in generation time, onset of puberty, and rate of spermatogenesis should all influence yearly mutation rates (1, 17).

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 10-fold, with the shortest generation time observed in prosimians [~3 y in galago and mouse lemurs (18)] and the longest generation time observed in humans (~29 y). 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 (19). Thus, even if the per cell division mutation rate remained constant, we should expect differences in yearly mutation rates among species.

Although 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 (6). For example, in mammals, CpG transitions show the least amount of variation in substitution rates among species (6). A plausible explanation is the source of mutations, as 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 the number of cell divisions (20–22).

In addition, even substitutions that have no effect on fitness may vary in their rate of accumulation among lineages because of biased gene conversion (BGC), the bias toward strong (S: G or C) rather than weak (W: A or T) bases that occurs in the repair of double-strand breaks (23). 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 (23, 24). The strength of BGC is a function of the degree of bias, the local recombination rate, and the effective population size of the species (23). The latter varies by three- to fourfold among primates (25), and the fine-scale recombination landscape is also likely to differ substantially across species (26).

Empirically, the extent to which substitution rates vary among primate lineages remains unclear. Kim et al. (27) compared two hominoids (human and chimpanzee) and two Old World Monkeys (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 ~31% between hominoids and OWMs, whereas rates of CpG transitions are almost identical (27). In turn, Elango et al. (28) found that the human branch is ~2% shorter than that in chimpanzee (considering the rates from the human–chimpanzee ancestor), and ~11% shorter than in gorilla (considering rates from the human–gorilla ancestor). Although these comparisons suggest 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 (29). We therefore revisit these questions using whole-genome sequence alignments of 10 primates, allowing for variable substitution rates along different lineages and explicitly modeling the context dependency of CpG substitutions.

## Results

We first estimate the number of autosomal substitutions on 10 primate lineages by applying Phylofit (30) to the Multiz sequence alignment (excluding gorilla and gibbon because of concerns about incomplete lineage sorting; *SI Appendix*, Note S1). This method allows us to estimate branch lengths, accounting for uncertainty in the ancestral reconstruction, recurrent substitutions at a site, and context-dependent effects of neighboring nucleotides at CpG sites (30).

To focus on putatively neutral sites in the genome, in which substitutions more faithfully reflect mutation patterns, we exclude conserved elements, coding exons, and transposable elements (referred to as CET in what follows; *SI Appendix*, Note S1). After filtering CET sites and removing missing data, we obtain ~562 Mb of whole-genome sequence alignment across 10 primates. In these filtered data, the total number of substitutions on the human lineage is similar to estimates in ancestral repeats (*SI Appendix*, Table S3), which are often considered a benchmark for strict neutral evolution (31), suggesting the substitutions we analyzed were largely neutral.

Across the 10 primate species, we find that the total substitution rates vary markedly (Fig. 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 2.63–2.74%), whereas rates on lineages leading to OWM are on average 3.57% (range: 3.55–3.59%), 1.33-fold higher. These findings are consistent with those of smaller studies (27). Similarly, when considering the distance from the hominoid–New World Monkey (NWM) ancestor, substitution rates leading to NWM are on average 6.92% (range: 6.89–6.94%), 1.64-fold higher than on the lineages leading to hominoids, which are on average 4.22% (range: 4.17–4.29%). Substitution rates are also 1.61-fold higher in lineages leading to bushbaby (a prosimian) compared with hominoids (Fig. 1). 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, however, and hence do not consider bushbaby in further analyses.

Using bootstrap resampling of 1-Mb regions of the genome suggests tiny SEs for the substitution rates (e.g., the SE on lineages leading from the hominoid–OWM ancestor to hominoids is

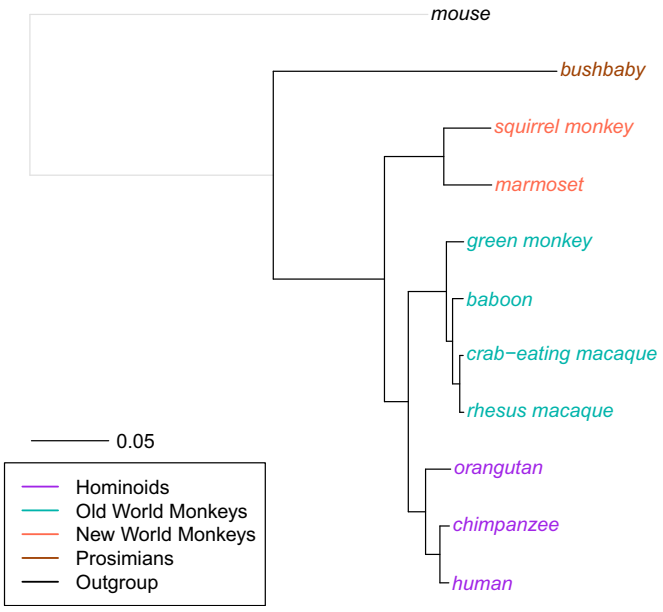



**Fig. 1.** Phylogenetic tree for 10 primates. Autosomal neutral substitution rates for 10 primates and an outgroup (mouse, shown in gray) from the Multiz dataset were estimated using Phylofit (see *SI Appendix*, Note S1 for details of dataset and filtering). Branch lengths reflect the expected number of neutral substitutions per site along each lineage. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/Figure1.R.

0.01%), as expected from such large datasets. These SEs are likely to be deceptively small, however, as the main source of uncertainty in our analysis is likely a result of 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 Enredo-Pecan-Ortheus (EPO) dataset (32)] and apply the same filters. When the species considered are matched between the two datasets, results are highly similar (*SI Appendix*, Note S3), and appear to be robust.

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 as a result of 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 as a result of replication errors. (Because CGI are often hypomethylated, we remove these regions from this analysis, and, unless specified otherwise, refer to transitions at CpG sites outside of CGIs as "CpG transitions.") Mathematical modeling of different substitution mechanisms predicts that mutations that are nonreplicative in origin and highly inefficiently repaired should depend on absolute time, rather than on the number of cell divisions, and hence should be more clocklike among species (21). In contrast, mutations that arise from replication errors, or are nonreplicative in origin but well repaired, should depend on the generation time and other life history traits, and therefore their substitution rates could vary considerably across primates (21, 33). Thus, a priori, we expect CpG transitions outside CGI to be more clocklike than other types of substitutions (assuming similar rates of deamination across species).

For our comparisons to not be confounded by biased gene conversion, we compare transitions at CpG sites with those occurring at 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 (*SI Appendix*, Fig. S2), they should, on average, be subject to similar strengths of biased gene conversion. Comparing hominoids 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 (Fig. 2). In contrast, when considering transitions at non-CpG G/C sites, there are on average 1.38-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 (Fig. 2). Thus, CpG transition rates are more similar across species, as observed in comparisons of smaller datasets of primates and mammals (6, 27). These results are robust to the choice of species of OWM and hominoids used; e.g., using gorilla instead of chimpanzee or gibbon instead of orangutan yields similar findings (*SI Appendix*, Figs. S3 and 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), and 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 (*SI Appendix*, Note S1), 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 (Fig. 3*A*). 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) (6) and 9 primates (0.15 Mb) (34). Interestingly, transitions at CpG sites inside CGI have a greater variance in substitution rates and behave similar to G/C transitions (Fig. 3*A*). The difference in behavior of CpGs inside and outside CGI is again consistent with the notion that when the source of mutation is primarily

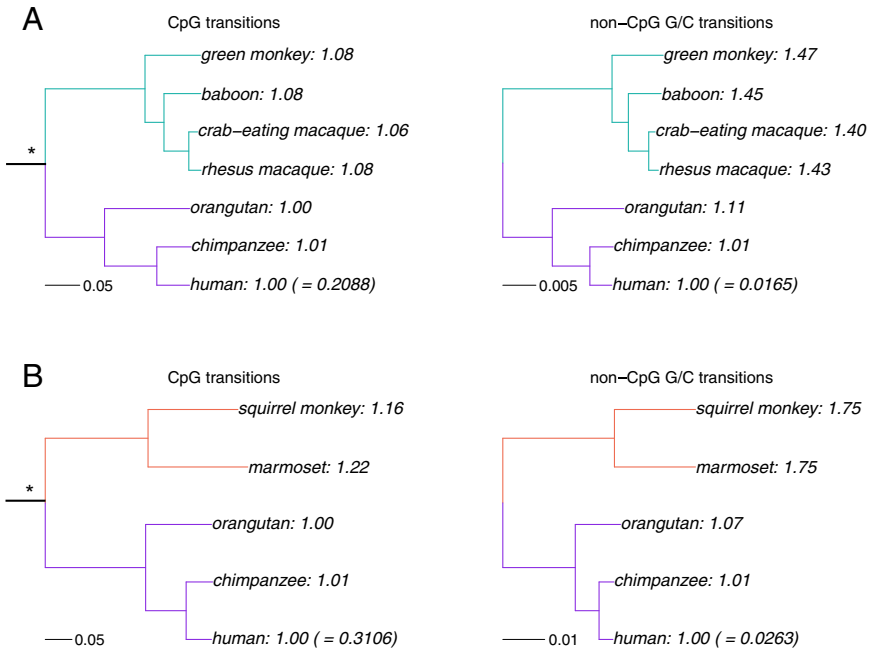



**Fig. 2.** Comparison of substitution rates in hominoids and monkeys. For transitions from CpG and non-CpG G/C sites, the total branch length is shown from either (*A*) the hominoid–OWM ancestor to each leaf, or (*B*) the hominoid–NWM ancestor to each leaf. The branch length from the root to the human tip was set to 1 (with the actual value in parenthesis), and other lineages normalized to the human branch length. Branches from root-hominoids are shown in purple, from root-OWM in green and from root-NWM in orange. *Hominoid–monkey (either OWM or NWM) ancestor used as root. R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/Figure2.R.

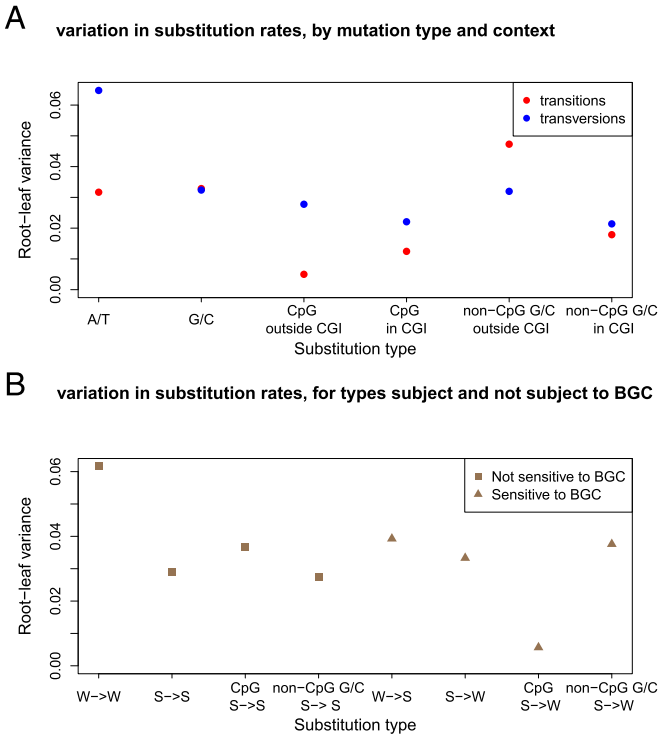



**Fig. 3.** Variance among lineages for different substitution types. (*A*) For each ancestral state and each context shown on the *x* axis, we estimate the total branch length from the root to each terminal leaf in the Multiz dataset as the inferred number of substitutions per site. We then calculate 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 [S (G/C) and 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 calculate the variance in the root-to-leaf distance across the nine primates used in *A*. Using only one species from each taxon yields similar results (not shown). R code to replicate this figure is available at: https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/Figure3.R.

nonreplicative, mutations may depend more on absolute time than numbers of cell divisions, whereas when they have sources that are dependent on the numbers of cell divisions, they will be 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 as a result of the effects of biased gene conversion, notably because of differences in effective population sizes (23). 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 (Fig. 3*B*; $P = 0.3$, based on a permutation test). Although this finding seems puzzling, given the three- to fourfold difference in effective population size of these species (25), it is consistent with results of Do et al., 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 twofold in their effective population sizes (35). 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 (human, chimpanzee, and gorilla). 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 (~30×) 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 (*SI Appendix*, Note S1). After removing missing data, nonneutral sites, and CGI, we obtain ~1.03 Gb of sequence for human–chimpanzee and ~1.02 Gb of sequence for 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, at 0.621% and 0.633%, respectively. This difference of 1.9% is, however, highly statistically significant, under the assumption of no systematic errors ($P < 10^{-20}$; *SI Appendix*, Note S1). When we consider the substitution rates at different mutation types, there are somewhat more pronounced differences for some types of substitutions, in inconsistent directions. For example, when comparing chimpanzee with 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 (*SI Appendix*, Fig. S9). Comparing human and gorilla lineages, differences in substitution rates are more pronounced: the gorilla branch (0.824%) is longer than the human (0.773%) branch by, on average, 6.6% ($P < 10^{-20}$; *SI Appendix*, Note S1). Again, different types of substitutions show distinct patterns, ranging between 0.96-fold at CpG transversions on the gorilla versus human branch to 1.10-fold for A/T transitions (*SI Appendix*, Fig. S10).

To check the reliability of these inferences, we also estimate the substitution rates using a second method based on a maximum-likelihood approach (36) (*SI Appendix*, Note S1). Although the absolute values of the substitution rates differ between the two methods, possibly as a result of methodological differences in calling ancestral states and assumptions about stationarity, the ratios of substitution rates between humans and chimpanzees (*SI Appendix*, Figs. S9 and S11) and between humans and gorillas (*SI Appendix*, Figs. S10 and S12) are almost identical. Although these results for the human–chimpanzee comparison match those obtained by Elango et al. (28), based on 75 Mb of data, our estimate of 1.07 for human–gorilla sequence difference is lower than the previous estimate of 1.11, based on 2 Mb of sequence data (28). Because 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 slightly too high.

Importantly, these observations imply that the mutation spectra, and not just the yearly mutation rate, are changing across primates. Notably, although the rate of substitutions involving CpG transitions is relatively stable across species, the proportion of substitutions involving CpG transitions varies across species. Beyond that, the substitution rates for other mutation types also vary considerably (*SI Appendix*, Fig. S13). 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 with Asians and Africans since these populations split (37).

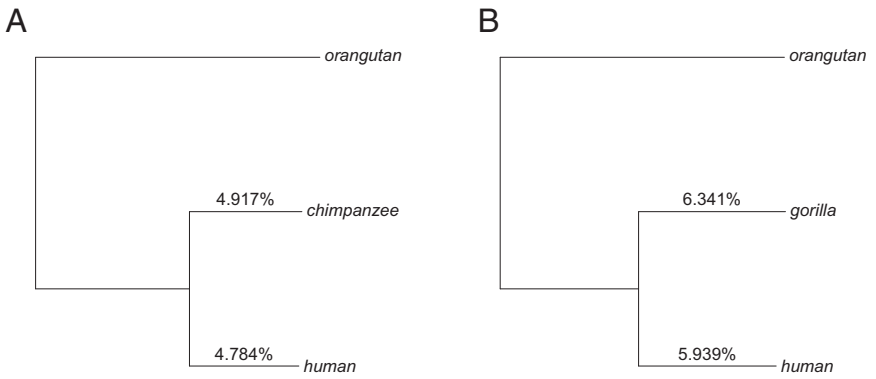



**Fig. 4.** Revised divergence time for hominines. We estimate the autosomal substitution rates for transitions at CpG sites by applying Phylofit to the high-coverage pairwise alignment of (*A*) human and chimpanzee and (*B*) human and gorilla. All hominines were were mapped to the orangutan reference genome. To infer divergence times, we use germline mutation rates for CpG transitions estimated from sequencing human pedigrees (see *SI Appendix*, Note S1 for details). We estimate average human and chimpanzee divergence time as 12.1 Mya, and average human and gorilla divergence time as 15.1 Mya. R code to replicate this figure is available at https://github.com/priyamoorjani/Molecular-clock_figures-and-data/blob/master/Figure4.R.

## Discussion

Evolutionary rates are faster in NWMs compared with OWMs, and in turn, rates in OWMs are faster than in humans and apes. These findings support the hominoid rate slowdown hypothesis (38, 39), 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 NWMs have a substantially shorter generation time (g = ∼6 y) than OWMs (g = ∼11 y), who in turn reproduce at younger ages than apes (g = ∼25 y; *SI Appendix*, Table S2). Within hominines, gorillas (g = ∼19 y) have a faster yearly rate than humans (g = ∼29 y) and chimpanzees (g = ∼25 y; *SI Appendix*, Table S2). To investigate whether the association between generation time and substitution rates is significant after controlling for the underlying phylogeny, we perform the phylogenetically independent contrast analysis (40) (*SI Appendix*, Note S1). Specifically, we assume the underlying phylogeny based on CpG transition rates (effectively assuming these are strictly clocklike) and then estimate the correlation between generation times and non-CpG substitution rates, controlling for the shared phylogenetic history. Using the nine species available for the analysis, the association is not significant ($r = 0.17$; $P = 0.7$), so the causal relationship remains to be established for primates.

An alternative approach is to ask whether differences in generation times and other life history traits can plausibly explain the variation in substitution rates. To this end, we use a model introduced by Amster and Sella (33) to describe mutations that are replicative in origin, which should also apply to mutations that are nonreplicative but well repaired (21). This model relates substitution rates to sex-specific life history and reproductive traits, and thus predicts how substitution rates are expected to differ among species. In applying the model, one option would be to examine the effect of one trait at a time. However, 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 ($r = 0.90$; $P = 0.002$) (*SI Appendix*, 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 OWMs and mutation parameters estimated from human pedigree studies (*SI Appendix*, Note S1 and Table S2). On that basis, we predict that yearly mutation rates in humans and chimpanzees should differ by ∼19%, and hominoids (using humans and chimpanzees here as not all parameter values are available for orangutans) and OWMs should differ by ∼86%. Thus, if anything, differences in life history traits in extant species predict even more variation in substitution rates than is observed (33).

The use of life history traits in extant species will exacerbate the expected differences in substitution rates if closely related species have had similar life histories throughout much of their evolutionary past. Fossil evidence suggests the age of puberty on the human lineage may have only recently increased, for example, and was lower in *Homo erectus* and Neanderthals (41, 42). Similar changes are likely to have occurred on the chimpanzee lineage as well. If we change the age of onset of puberty in humans to 9 y, the difference in rates between humans and chimpanzees is only expected to be ∼5%. One implication, then, of finding such similar substitution rates in humans and chimpanzees is that their life histories may have been fairly 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. One way to overcome this difficulty is to explicitly model the changes in life history traits within species and over the course of primate evolution (33). Taking this approach, Amster and Sella (33) show that accounting for variation in generation time, age of onset of puberty, and rate of spermatogenesis in extant apes helps to reconcile the split times estimated on the basis of molecular and fossil evidence (33). Their method, however, requires knowledge of life history traits in both extant and ancestral populations.

An alternative is 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 nonzero, possibly because a subset of these mutations occurs due to replication errors, or because repair is not completely inefficient, or mutations do not accumulate in strict proportion to parental ages (21). Nonetheless, CpG transitions appear to be least affected by life history differences across species, accumulating in a quasi-clocklike manner. Moreover, in humans, they contribute almost a fifth of all de novo mutations, and so provide enough data for precise estimation (10).

With these considerations in mind, we reestimate the divergence and split times of humans, chimpanzees, and gorillas, using substitution rates estimated only at CpG transitions. Assuming the per year mutation rate for CpG transitions obtained in ref. 10 (*SI Appendix*, Note S1), we estimate that humans diverged from chimpanzees ∼12.1 Mya and from gorillas ∼15.1 Mya (Fig. 4). Assuming further that the effective population size of the human–ape ancestor was five times the current population size (as estimated by refs. 43, 44), the human–chimpanzee split time is ∼7.9 Mya, and

the human–gorilla split time is 10.8 Mya. We note that there is substantial uncertainty in estimates of ancestral population size of apes, with previous estimates ranging between 50,000 and 100,000 (43–45). Accounting for this uncertainty provides estimates of human–chimpanzee split time in the range of 6.5–9.3 Mya, and human–gorilla split time in the range of 9.4–12.2 Mya. Reassuringly, these estimates are similar to those obtained by explicitly modeling the dependence of replicative mutations on life history traits in hominines (33). Moreover, they are in broad agreement with evidence from the fossil record, which suggests a human–chimpanzee split time of 6–10 Mya and a human–gorilla split time of 7–12 Mya (46–51). Thus, within hominines, there is no obvious discrepancy between phylogenetic and pedigree-based estimates of mutation rates, once the effect of life history traits on mutation rates is taken into account (33).

## Materials and Methods

We used Phylofit (30) to estimate autosomal substitution rates for different mutation types, using the following three datasets: a 12-primate whole-genome sequence alignment, with mouse as an outgroup, that is part of a 100-way mammalian phylogeny, mapped using Multiz (52) (referred to as the Multiz dataset); a seven-primate whole-genome alignment, mapped using the Enredo-Pecan-Ortheus pipeline (32) (referred to the EPO dataset); and high coverage genomes for a human (of European descent) that we sequenced (*SI Appendix*, Note S2), a chimpanzee (Ind-D from ref. 11), and a gorilla [Delphi from ref. 44; data kindly provided by Tomas Marques-Bonet, Institut Biologia Evolutiva, Universitat Pompeu Fabra/Spanish National Research Council (CSIC) (referred to as the high-coverage hominoid dataset)]. These genomes were mapped to the orangutan reference genome (ponAbe2) (53), which should be equidistant to humans and extant African great apes (assuming no variation in substitution rates). We matched these species for coverage, alignment, and mapping pipelines to minimize the effects of technical artifacts. For details, see *SI Appendix*, Note S1.

**ACKNOWLEDGMENTS.** We thank David Pilbeam, Guy Amster, Guy Sella, Jenny Tung, Minyoung Wyman, Susan Alberts, and Ziyue Gao for helpful discussions. We thank Nick Patterson and Heng Li for technical advice for mapping and alignment of high-coverage genomes and Melissa Hubisz and Adam Siepel for advice on running Phylofit. P.M. was supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award F32 GM115006-01. C.E.G.A. was supported by a Science Without Borders fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico Brazil (PDE 201145/2015-4). The computing in this project was supported by two National Institutes of Health instrumentation grants (S10OD012351 and S10OD021764) received by the Department of Systems Biology at Columbia University.

1. Ségurel L, Wyman MJ, Przeworski M (2014) Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 15:47–70.
2. Kimura M (1984) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, New York).
3. Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, eds Bryson V, Vogel HJ (Academic, New York), pp 97–166.
4. Kumar S (2005) Molecular clocks: Four decades of evolution. *Nat Rev Genet* 6(8): 654–662.
5. Wilson Sayres MA, 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* 65(10):2800–2815.
6. Hwang DG, Green P (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci USA* 101(39):13994–14001.
7. Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231(4744):1393–1398.
8. Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci USA* 82(6):1741–1745.
9. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128(2):415–423.
10. 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.
11. Venn O, et al. (2014) Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* 344(6189):1272–1275.
12. Scally A, Durbin R (2012) Revising the human mutation rate: Implications for understanding human evolution. *Nat Rev Genet* 13(10):745–753.
13. Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA* 90(9):4087–4091.
14. 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 Evol Biol* 7(1):95.
15. Wong A (2014) Covariance between testes size and substitution rates in primates. *Mol Biol Evol* 31(6):1432–1436.
16. Drost JB, Lee WR (1995) Biological basis of germline mutation: Comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ Mol Mutagen* 25(Suppl 26):48–64.
17. Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1(1):40–47.
18. Gage TB (1998) The comparative demography of primates: With some comments on the evolution of life histories. *Annu Rev Anthropol* 27:197–221.
19. Ramm SA, Stockley P (2009) Sperm competition and sperm length influence the rate of mammalian spermatogenesis. *Biol Lett* 6(2):219–221.
20. Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* 287(5782):560–561.
21. Gao Z, Wyman MJ, Sella G, Przeworski M (2016) Interpreting the dependence of mutation rates on age and time. *PLoS Biol* 14(1):e1002355.
22. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* 274(5673):775–780.
23. Duret L, Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 10:285–311.
24. 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.
25. Leffler EM, et al. (2012) Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol* 10(9):e1001388.
26. Stevison LS, et al.; Great Ape Genome Project (2016) The Time Scale of Recombination Rate Evolution in Great Apes. *Mol Biol Evol* 33(4):928–945.
27. Kim S-H, Elango N, Warden C, Vigoda E, Yi SV (2006) Heterogeneous genomic molecular clocks in primates. *PLoS Genet* 2(10):e163.
28. Elango N, Thomas JW, Yi SV; NISC Comparative Sequencing Program (2006) Variable molecular clocks in hominoids. *Proc Natl Acad Sci USA* 103(5):1370–1375.
29. 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.
30. Siepel A, Haussler D (2004) Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol* 21(3):468–488.
31. Ananda G, Chiaromonte F, Makova KD (2011) A genome-wide view of mutation rate co-variation using multivariate analyses. *Genome Biol* 12(3):R27.
32. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E (2008) Enredo and Pecan: Genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* 18(11):1814–1828.
33. Amster G, Sella G (2016) Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc Natl Acad Sci USA* 113(6):1588–1593.
34. Lee H-J, Rodrigue N, Thorne JL (2015) Relaxing the Molecular Clock to Different Degrees for Different Substitution Types. *Mol Biol Evol* 32(8):1948–1961.
35. Do R, et al. (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47(2):126–131.
36. Duret L, Arndt PF (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* 4(5):e1000071.
37. Harris K (2015) Evidence for recent, population-specific evolution of the human mutation rate. *Proc Natl Acad Sci USA* 112(11):3439–3444.
38. Goodman M (1963) Serological analysis of the systematics of recent hominoids. *Hum Biol* 35:377–436.
39. Yi SV (2013) Morris Goodman's hominoid rate slowdown: The importance of being neutral. *Mol Phylogenet Evol* 66(2):569–574.
40. Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125(1):1–15.
41. Graves RR, Lupo AC, McCarthy RC, Wescott DJ, Cunningham DL (2010) Just how strapping was KNM-WT 15000? *J Hum Evol* 59(5):542–554.
42. Gluckman PD, Hanson MA (2006) Changing times: The evolution of puberty. *Mol Cell Endocrinol* 254-255:26–31.
43. Wall JD (2003) Estimating ancestral population sizes and divergence times. *Genetics* 163(1):395–404.
44. Prado-Martinez J, et al. (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471–475.
45. Scally A, et al. (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483(7388):169–175.
46. Wilkinson RD, et al. (2011) Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst Biol* 60(1):16–31.
47. Begun DR (2015) Fossil record of Miocene hominoids. *Handbook of Paleoanthropology*, eds Henke W, Tattersall I (Springer, Berlin), pp 1261–1332.
48. Brunet M, et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418(6894):145–151.
49. 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.
50. White TD, et al. (2009) *Ardipithecus ramidus* and the paleobiology of early hominids. *Science* 326(5949):64–86.
51. Wood B, Harrison T (2011) The evolutionary context of the first hominins. *Nature* 470: 347–352.
52. Karolchik D, et al. (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 42(Database issue, D1):D764–D770.
53. Locke DP, et al. (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469(7331):529–533.

# 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 ($\mu_y$) 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 $\mu_F$ 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 $D_M/\tau$ 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

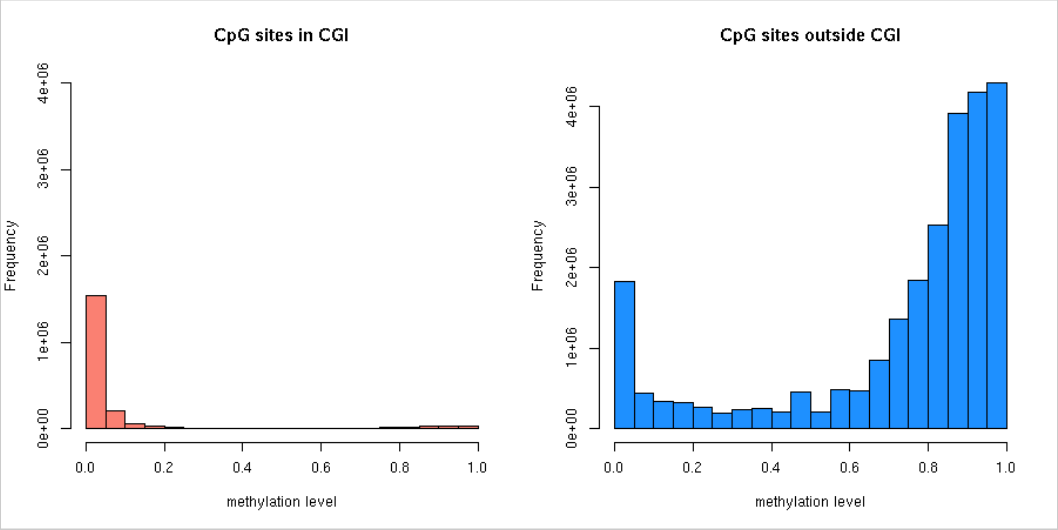


**Figure S2: Distribution of CpG and non-CpG G/C sites across the human genome.** The proportion of CpG and non-CpG G/C sites in the human genome, as a function of the recombination rate is shown. After filtering non-neutral sites and CGI (see Note S1) in the Multiz dataset, the proportions of CpG and 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

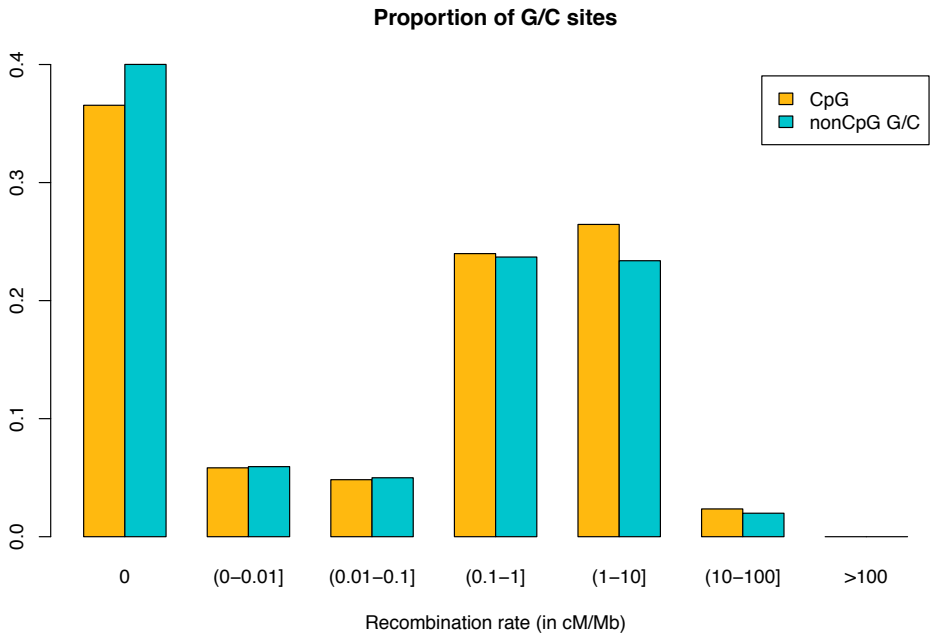


**Figure S3: Comparison of substitution rates in hominoids and Old World Monkeys (OWMs) using alternate topologies.** Due to concerns about the possible effects of incomplete lineage sorting, we analyzed gorilla and chimpanzee (and gibbon and orangutan) separately. Each sub-figure shows a different set 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

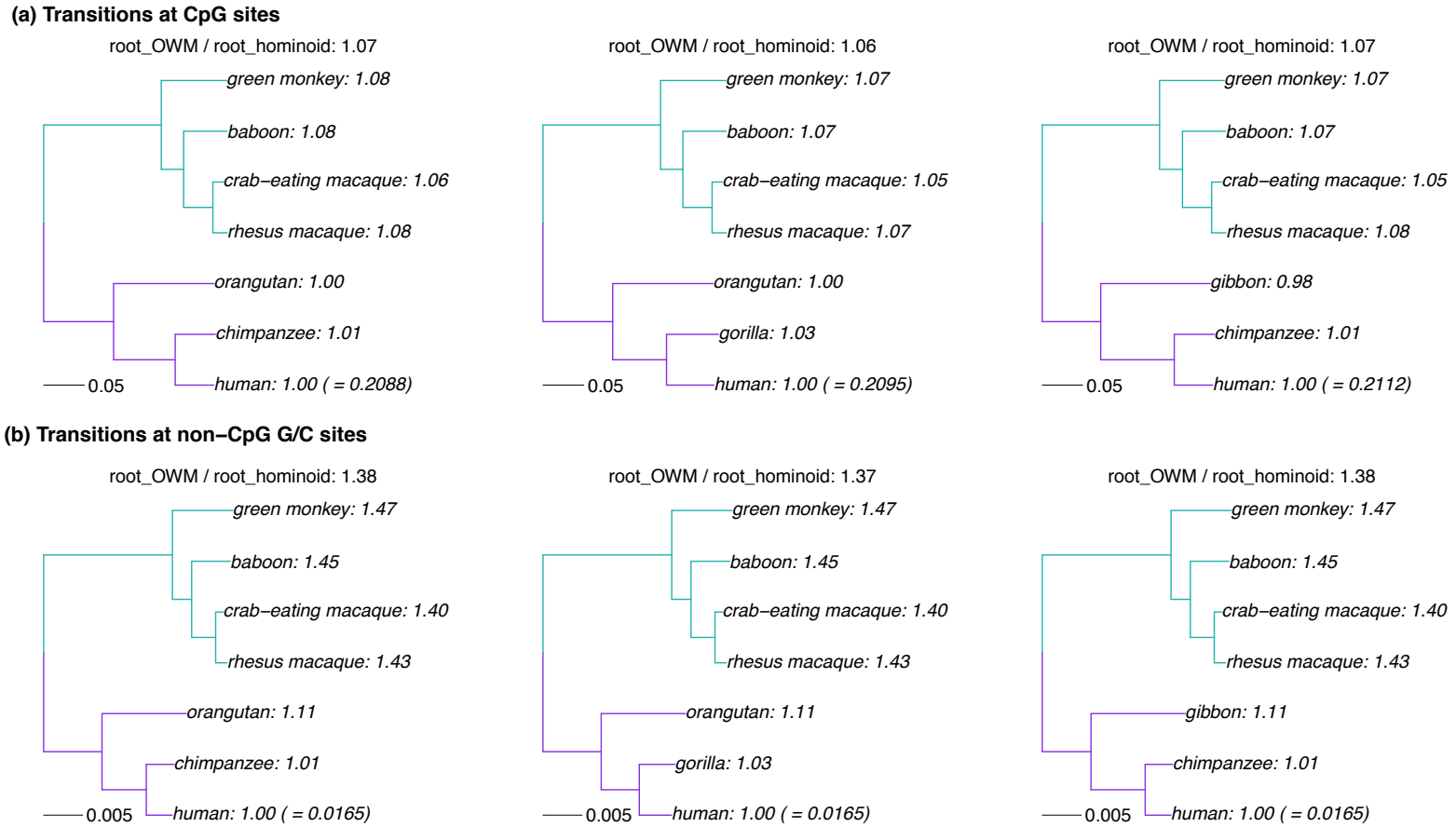


**Figure S4: Comparison of substitution rates in hominoids and New World Monkeys (NWMs) using alternate topologies.** Due to concerns about the possible effects of incomplete lineage sorting, we analyzed gorilla and chimpanzee (and gibbon and orangutan) separately. Each sub-figure shows a different set 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

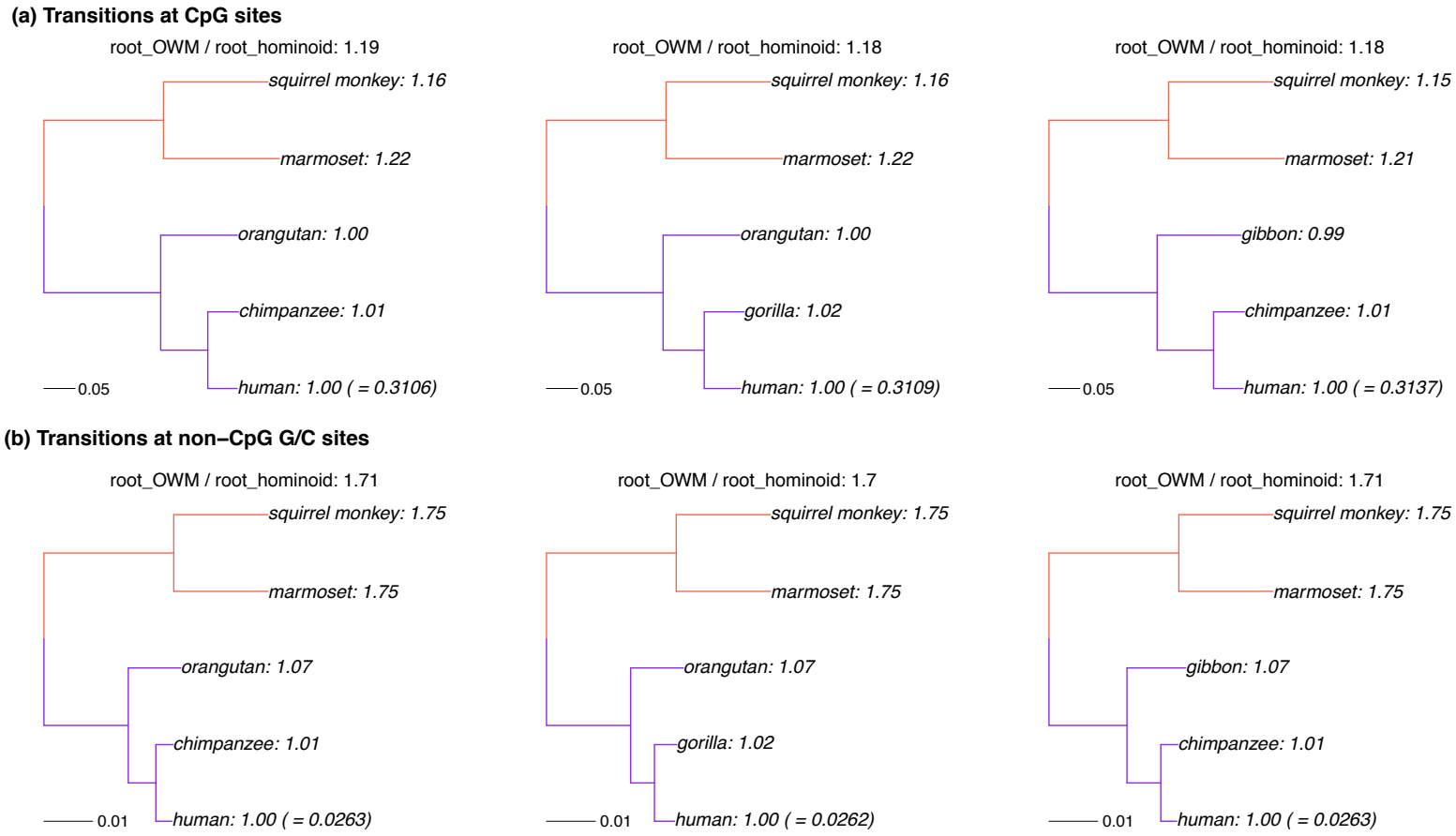


**Figure S5: Phylogenetic tree for the six primates in EPO dataset.** We estimated neutral substitution rates for six primates from the EPO dataset using Phylofit (see Note S1 for details). Branch lengths reflect the expected number of neutral substitutions per site along each lineage. We excluded gorilla 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

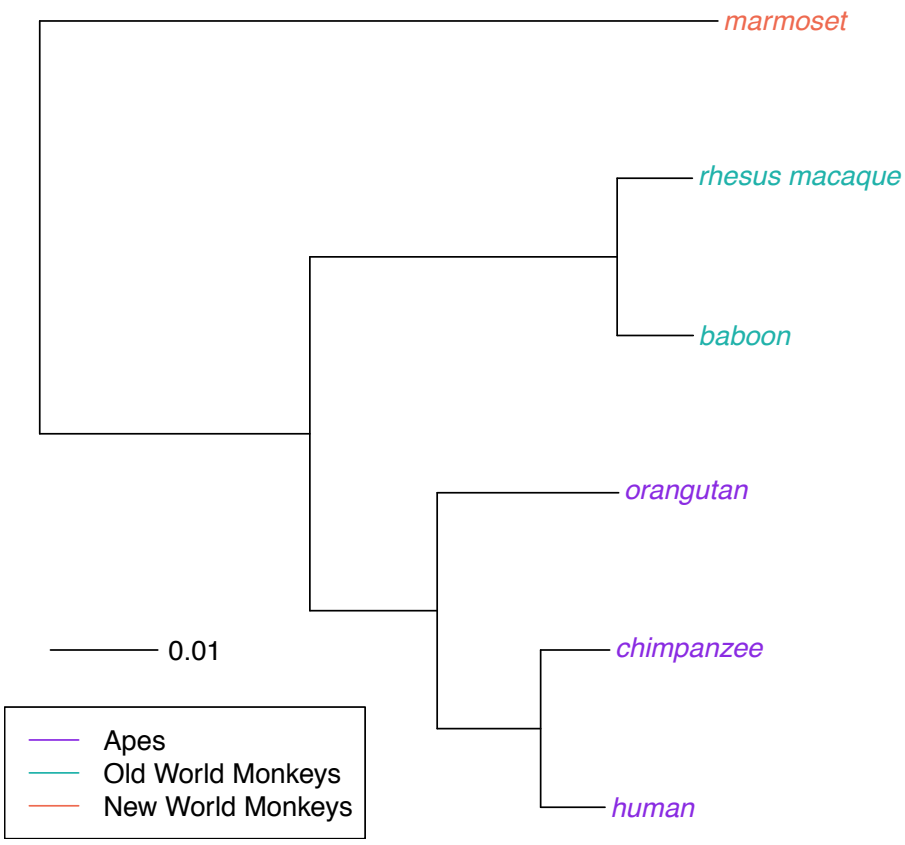


**Figure S6: Comparison of substitution rates in hominoids and OWMs using different datasets.** For each dataset (Multiz or EPO), we estimated the total branch length from the hominoid-OWM ancestor (root) to each leaf. The branch length from the root to the human tip was set to 1 (the actual value is 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

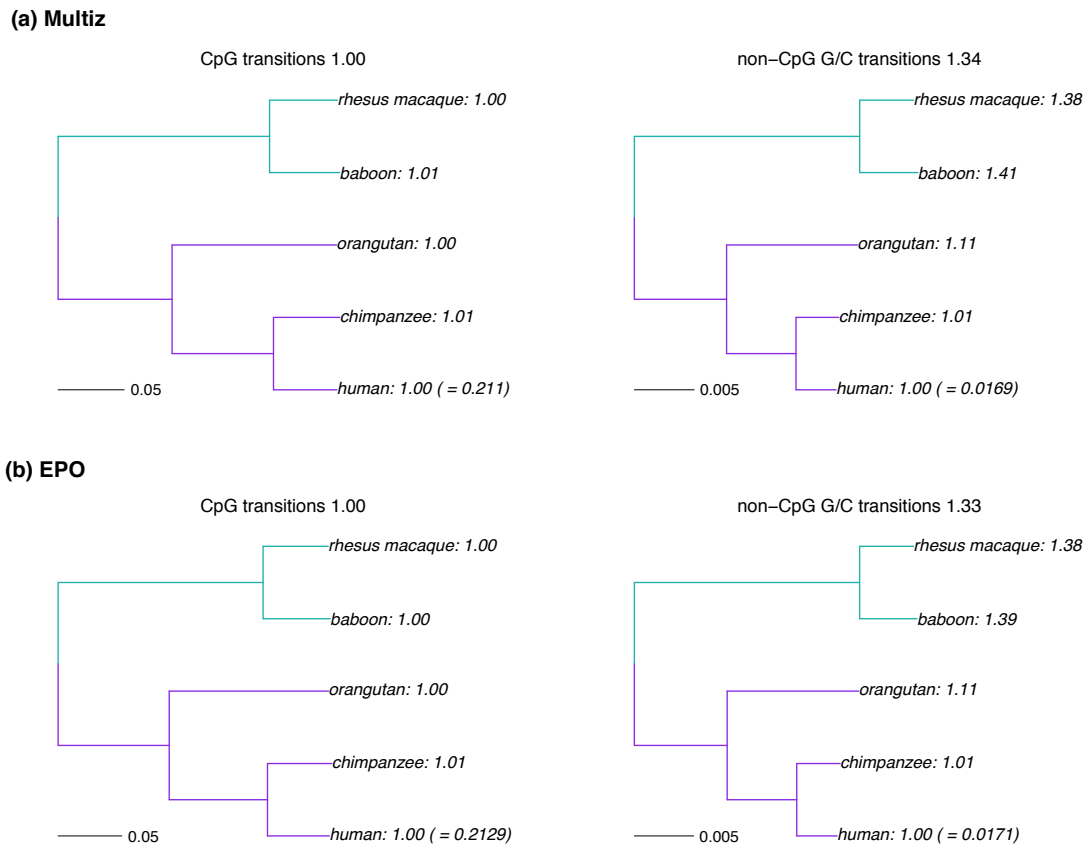
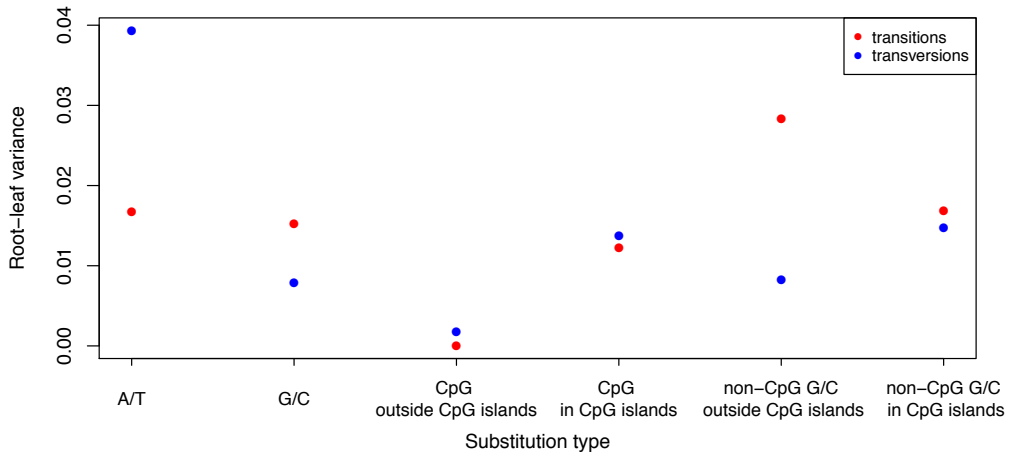


**Figure S7: Variance among lineages for distinct substitution types, estimated from different datasets.** For each ancestral state and each context shown on the x-axis, we estimated the total branch length from the root to each terminal leaf as the inferred number of substitutions per site, in (a) Multiz 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**

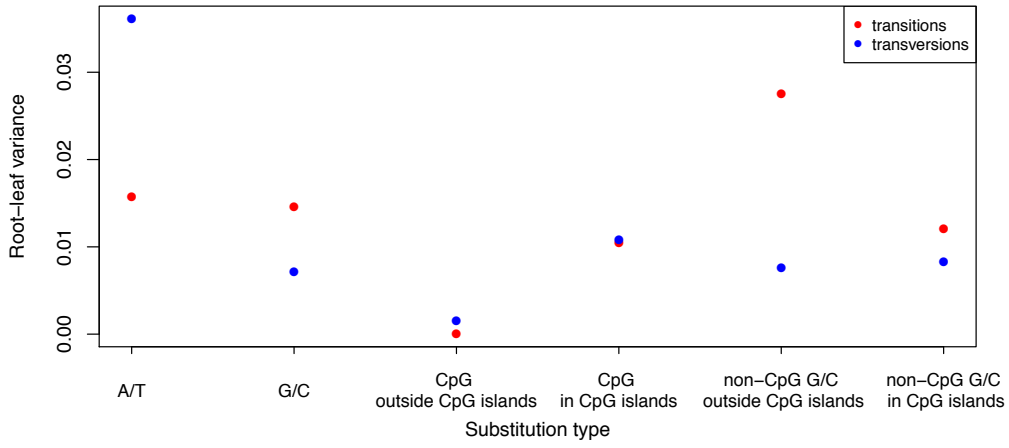
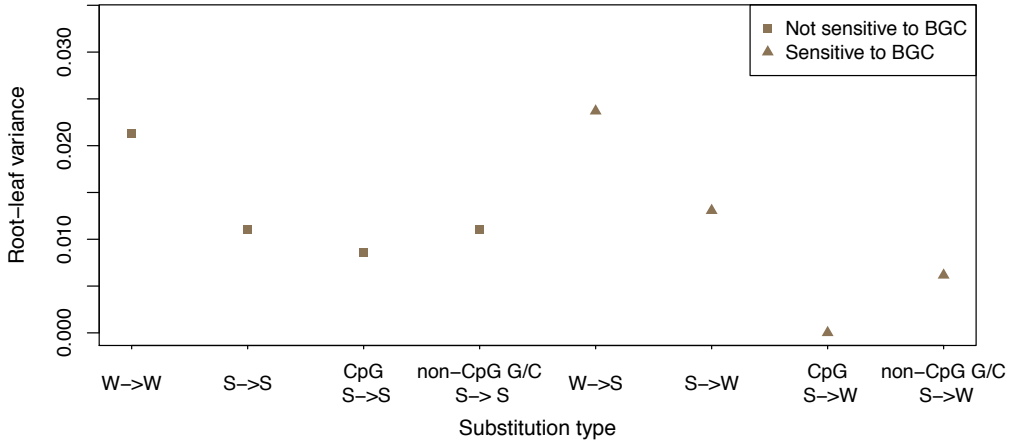


**Figure S8: Effect of biased gene conversion across lineages estimated for different datasets.** For each substitution type (strong (S; G/C) and weak (W; A/T)) and each ancestral context shown on the x-axis, we estimated the total branch length from the root to each terminal leaf as the inferred number 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**

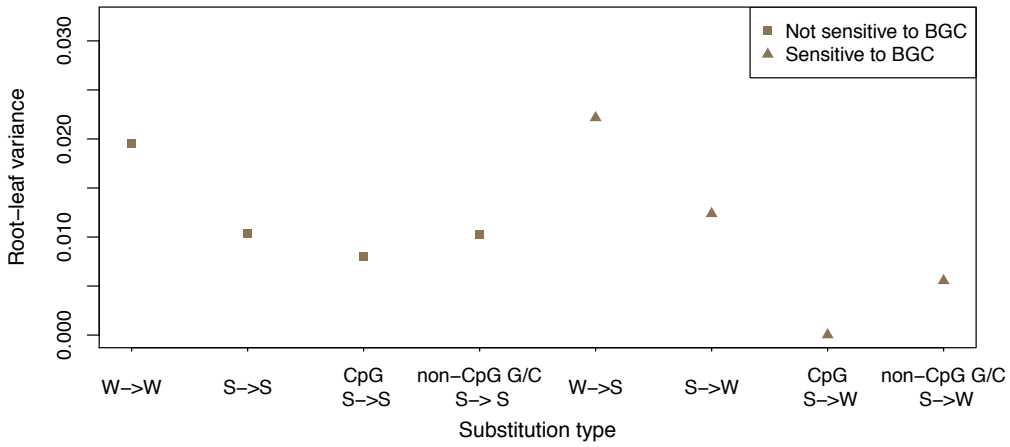


**Figure S9: Comparison of substitution rates in human and chimpanzee using Phylofit.** For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio of the substitution rate 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

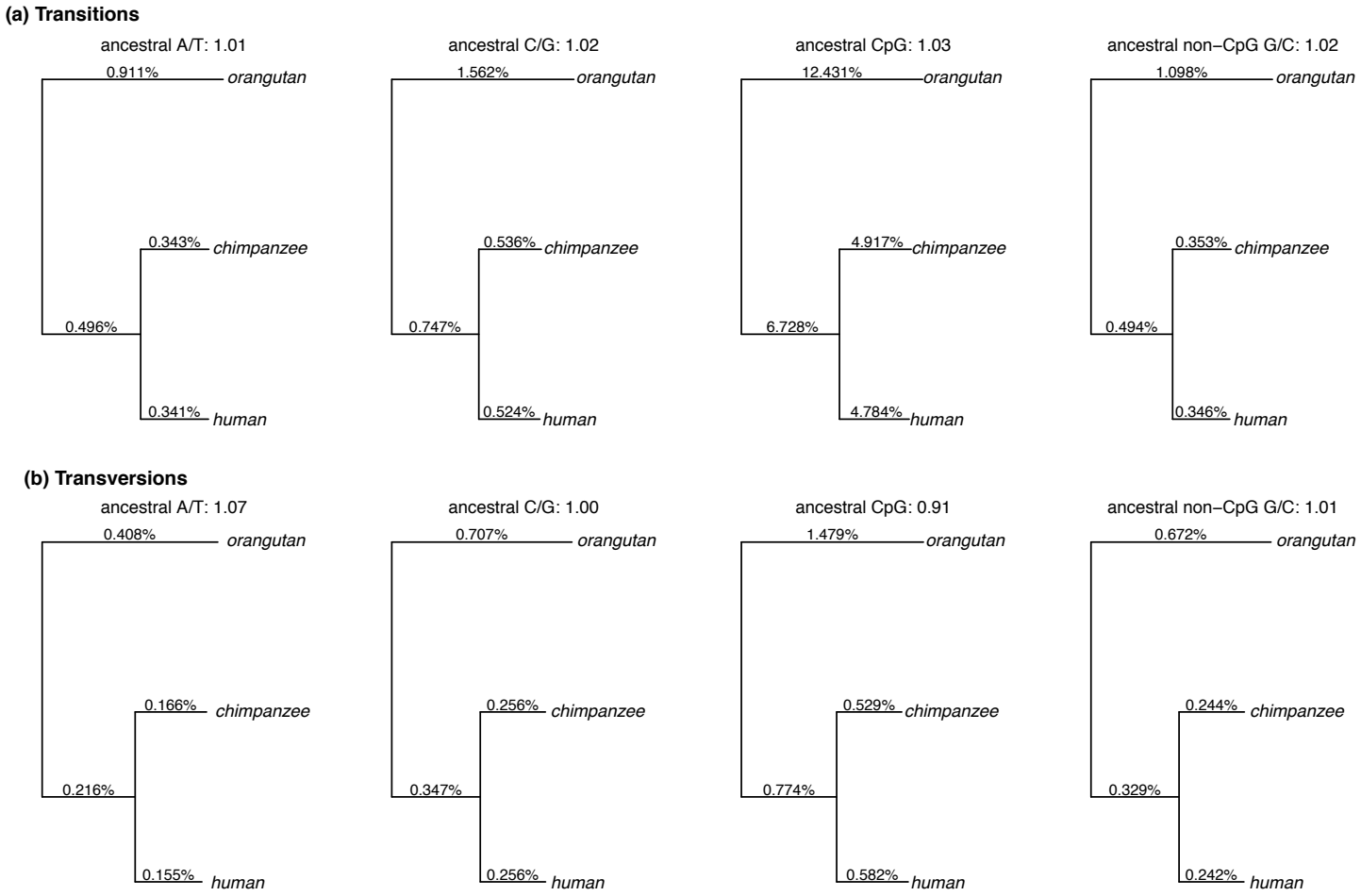


**Figure S10: Comparison of substitution rates in human and gorilla using Phylofit.** For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of the substitution rate in 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

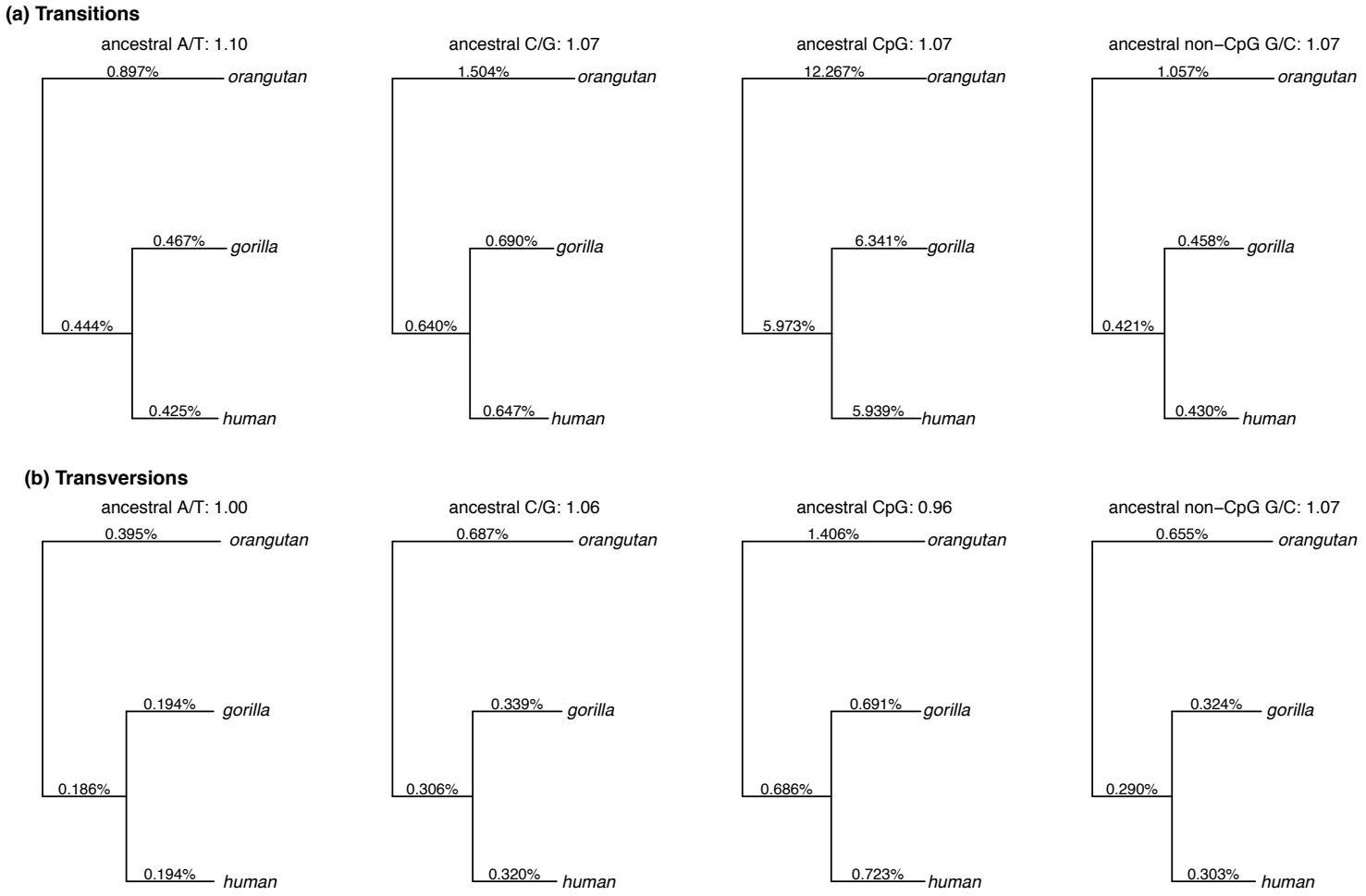


**Figure S11: Comparison of substitution rates in human and chimpanzee using the maximum likelihood approach.** For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and chimpanzee mapped to the orangutan reference genome. The ratio 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

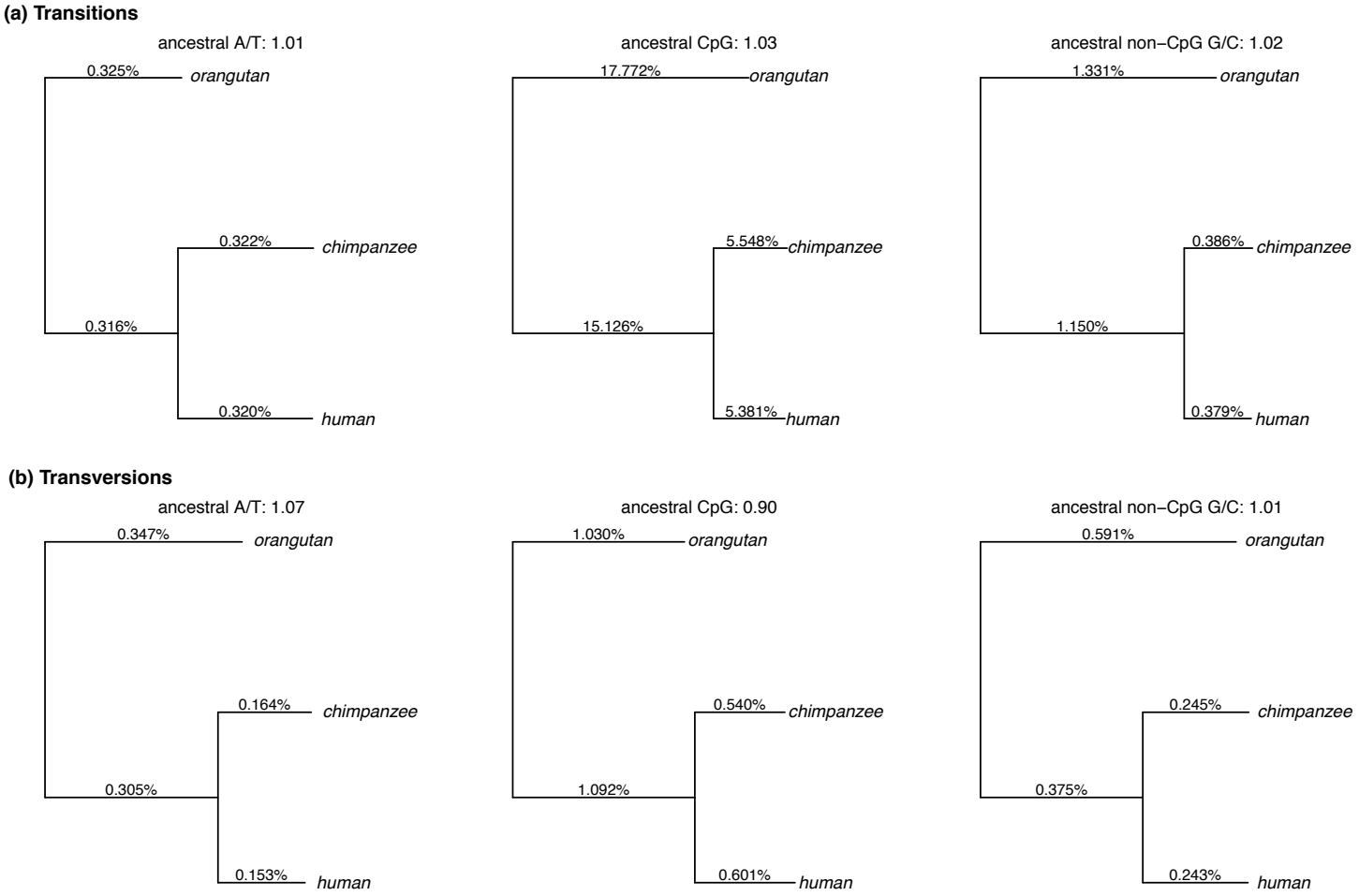


**Figure S12: Comparison of substitution rates in human and gorilla using the maximum likelihood approach.** For each substitution type, we estimated the autosomal substitution rate using the high coverage pairwise alignment of human and gorilla mapped to the orangutan reference genome. The ratio of 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

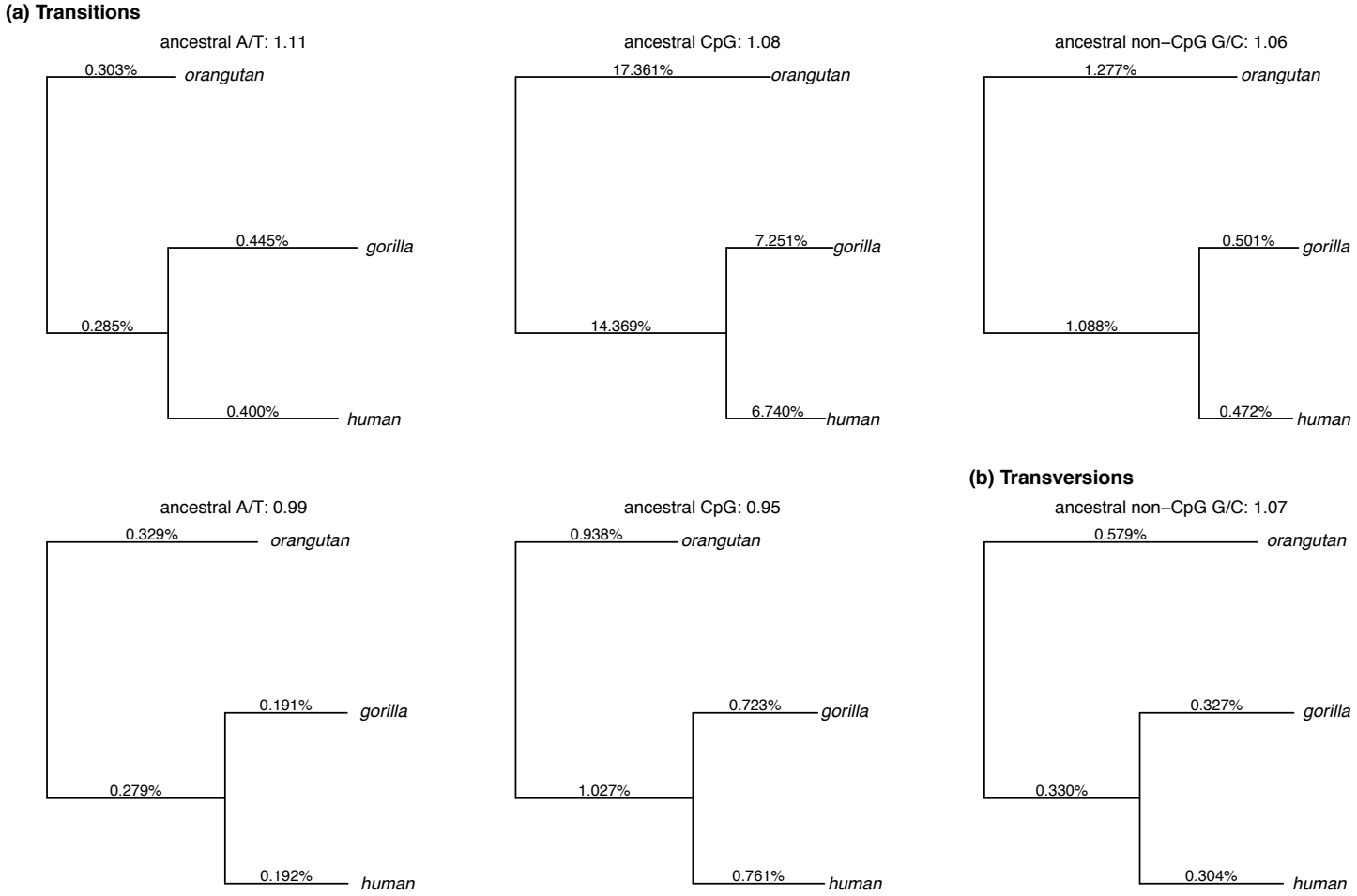


**Figure S13: Mutation spectrum across primates.** We estimated the number of substitutions along each lineage for each mutation type. We then normalized the number of substitutions of a given type to the number of transitions from ancestrally CpG sites that occurred on that lineage. R code to replicate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

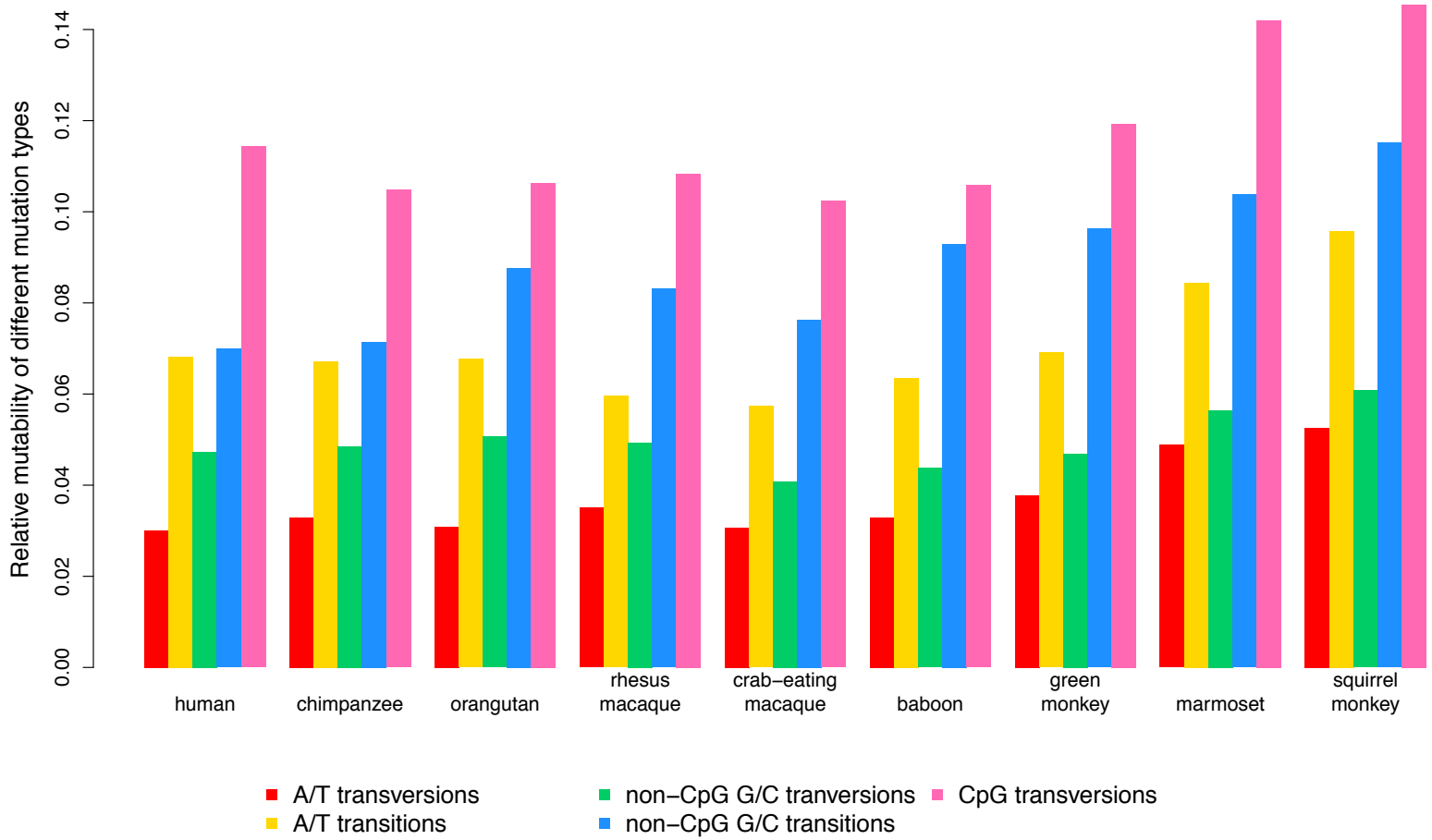


**Table S1. Online source of annotation for transposable elements, coding exons, CpG Islands (CGI), and conserved sites.**

| 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.