Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case

Mario dos Reis¹,², Gregg F. Gunnell³*, José Barba-Montoya², Alex Wilkins³,⁴, Ziheng Yang² and Anne D. Yoder⁵.

Correspondence to: m.dosreisbarros@qmul.ac.uk

1. School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK.
2. Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK.
3. Division of Fossil Primates, Duke Lemur Center, Durham, NC USA.
4. Department of Anthropology, The Ohio State University, Columbus, OH USA.
5. Department of Biology, Duke University, Durham, NC USA.

*Deceased; we dedicate this work to his memory and to the lasting value of his work
Abstract.—Primates have long been a test case for the development of phylogenetic methods for divergence time estimation. Despite a large number of studies, however, the timing of origination of crown Primates relative to the K-Pg boundary and the timing of diversification of the main crown groups remain controversial. Here we analysed a dataset of 372 taxa (367 Primates and 5 outgroups, 3.4 million aligned base pairs) that includes nine primate genomes. We systematically explore the effect of different interpretations of fossil calibrations and molecular clock models on primate divergence time estimates. We find that even small differences in the construction of fossil calibrations can have a noticeable impact on estimated divergence times, especially for the oldest nodes in the tree. Notably, choice of molecular rate model (auto-correlated or independently distributed rates) has an especially strong effect on estimated times, with the independent rates model producing considerably more ancient age estimates for the deeper nodes in the phylogeny. We implement thermodynamic integration, combined with Gaussian quadrature, in the program MCMCTree, and use it to calculate Bayes factors for clock models. Bayesian model selection indicates that the auto-correlated rates model fits the primate data substantially better, and we conclude that time estimates under this model should be preferred. We show that for eight core nodes in the phylogeny, uncertainty in time estimates is close to the theoretical limit imposed by fossil uncertainties. Thus, these estimates are unlikely to be improved by collecting additional molecular sequence data. All analyses place the origin of Primates close to the K-Pg boundary, either in the Cretaceous or straddling the boundary into the Palaeogene.

[Primates; phylogenomic analysis; molecular dating; relaxed clock; fossil; Bayesian analysis; Bayes factors]
Divergence time estimation is fundamentally important to every field of evolutionary analysis. Reliable estimates of the timing of speciation events across the Tree of Life allow us to correlate these events with both biotic and abiotic phenomena on geological and more recent timescales, thus illuminating those that are most closely associated with periods of rapid diversification (Zhou et al. 2012; Andersen et al. 2015), evolutionary stasis (Alfaro et al. 2009, Liu et al. 2014), or high rates of extinction (Prum et al. 2015; Zhang et al. 2016). The prospect of employing genomic information for discovering the geological age of the Primates began virtually simultaneously with the publication of Zuckerkandl and Pauling’s (1965) molecular clock hypothesis. Sarich and Wilson (1967) employed a strict clock interpretation of immunological distance data to hypothesize that humans and other African apes (i.e., chimp and gorilla) shared a common ancestor as recently as five million years ago. This was revolutionary at the time given the implications for the necessarily rapid evolution of bipedal locomotion in the hominin lineage, and accordingly, drew considerable attention from the anthropological community (Read and Lestrel, 1970; Uzzell and Pilbeam, 1971; Lovejoy et al., 1972; Radinsky, 1978; Corruccini et al, 1980). Despite this interest, it wasn’t until the 1980s that the field of divergence time estimation assumed a relatively modern flavour. It was then that investigators began to apply statistical models to DNA sequence data for the purposes of branch length and divergence time estimation (e.g., Hasegawa et al., 1985). Remarkably, these studies first emerged at a time when the sister-lineage relationship of humans to chimps was considered highly controversial (e.g., Goodman et al., 1983) — a relationship that is now considered unequivocal. Though it was Zuckerkandl and Pauling’s (1965) "molecular clock" hypothesis that made divergence time estimation feasible, it became clear very soon thereafter that there are myriad violations to a uniform clock. Thus, in subsequent decades, increasingly sophisticated models have been developed for relaxing the assumptions of a strict clock (Thorne et al. 1998; Kishino et al. 2001; Thorne, and Kishino 2002; Drummond et al. 2006; Lepage, et al. 2007; Rannala and Yang 2007, Heath et al. 2012). These models can be loosely divided into two categories: autocorrelated models, wherein rates of evolution in daughter species are statistically distributed around the parental rates (Sanderson 1997; Thorne et al. 1998; Aris-Brosou and Yang 2002), and
uncorrelated models, wherein each lineage on the tree is free to assume a fully
independent rate (Drummond et al. 2006; Rannala and Yang 2007; Paradis 2013).

A parallel challenge to divergence time analysis can be observed in the development of
calibration strategies (Marshall 1990; Yang and Rannala 2006; Benton and Donoghue
2007; Marshall 2008; Dornburg et al. 2011). Given that branch lengths on a phylogeny
are the product of rate and time, investigators must assume one to infer the other. The
most typical method for breaking this impasse is to employ fossil data as calibrations for
one or more nodes in a given phylogeny so that the ages of other nodes can be inferred
using DNA sequence data. This places an enormous burden on both the correct
placement and age assignment of the fossils. If they are misplaced (i.e., assigned to
clades where they do not belong) or if their geological ages are misinterpreted,
divergence time estimates for the entire tree can be severely compromised (Martin
1993). The uncertainty imposed by paleontological ambiguity has not been as widely
appreciated as have been the uncertainties introduced by the finite amount of DNA
sequence data, which with the "genomics revolution," is becoming steadily less
problematic.

We have reached a state of the art where branch lengths can be estimated with very high
precision. The combination of genome-scale data, sophisticated molecular evolutionary
models, and ever-increasing computational power has brought us to this point.
Advances in DNA sequencing technology are such that virtually every major clade has at
least a few species represented by whole-genome sequences, and this trend is rapidly
accelerating. Bayesian methods have been developed such that parameter space can be
explored during MCMC estimation, and though violations of the molecular clock will
continue to present problems, methods for measuring and accommodating rate
variation across phylogenies are explicit and generalizable. And finally, the computing
power to integrate this information is increasing steadily. But because of the
confounding effect of non-independence of rate and time, the expectation of a
conventional Bayesian analysis –that infinite data will eventually overcome prior
information and provide posterior distributions with certainty– cannot be met (dos Reis
and Yang 2013; Zhu et al. 2015).
Thus, the field at present is focused on developing a better understanding of the effects of relaxed clock model choice, and on the impacts of calibration points, both with regard to abundance and placement in the phylogeny. Furthermore, in addressing these challenges, it is an open question as to whether simulation studies or tests of empirical data will be more informative for our understanding of best practices. With regard to clock model choice, an empirical study of three independent datasets showed that autocorrelated models outperform uncorrelated models, though the same study found a "high sensitivity" to the relaxation model employed (Lepage et al. 2007), while another empirical study found, however, that an independent rates model was superior (Linder et al. 2011). Simulation studies have only recently been employed, finding that even with complete taxon sampling, rate autocorrelation is challenging to detect (Ho et al. 2015). This has led to the conclusion that rigorous model selection should be conducted on a case-by-case basis, utilizing a wide range of real data sets, and thus comprising a promising avenue for future research (Ho et al. 2005; Ho and Duchene 2014; Ho et al. 2015).

With regard to fossil calibration strategies, simulation studies (e.g., Duchene et al. 2014) have thus far agreed with previous empirical studies in finding that multiple calibrations are fundamentally important to accurate age estimation (Soltis et al. 2002; Yoder and Yang 2004; Linder et al. 2005; Rutschmann et al. 2007; Marshall 2008). Duchene et al. (2014) noted that calibrations close to the root of the phylogeny are most powerful. They also found that a significant source of error in divergence time estimation relates to misspecification of the clock model, especially when there are few informative calibrations. We cannot stress enough how sensitive posterior time estimates are to fossil information in constructing the time prior (Inoue et al. 2010). For example, different fossil calibration strategies have led to substantially different estimates of divergence times for mammals (dos Reis et al., 2014a) and early animals (dos Reis et al. 2015), regardless of how precise the branch length estimates are (Warnock et al. 2015).

Thus, the field has reached the stage wherein there is general agreement that the choice of clock model and calibration strategy are fundamentally important to the accuracy of resulting age estimates, and thus, the way forward will clearly involve both empirical and simulation approaches to the problem.
Here, we hope to contribute to this progress by conducting an exploration of model choice and calibration strategy in a classic empirical system: the Primates. Despite the fact that it is a relatively small and biologically uniform clade, primates have been inordinately and repeatedly the subject of divergence time analysis, with the first studies appearing at the very outset of molecular clock studies (Sarich and Wilson 1967), up to phylogenomic studies encompassing a large set of primate species (Perelman et al. 2011; Springer et al. 2012). This is largely due, undoubtedly, to the fact that we ourselves are members of this clade and can thus be forgiven for a persistent curiosity about our ancestral history. Age estimates for major primate divergence events have varied broadly among different studies (see Table 1), though one result has been relatively constant throughout: primate origins have been typically shown to predate the Cretaceous-Paleogene (K-Pg) mass-extinction event.

Our study explores the effects of an autocorrelated versus an uncorrelated rate model on age estimates, and also explores the consequences of two different interpretations of both the age and the placement of key fossils with the living primate radiation. We apply these two strategies to a large phylogenomic dataset for Primates (372 species and 3.4 million aligned base pairs). Until very recently, reliable calculations of branch lengths and age estimates within an analysis of this magnitude would have been beyond the capacity of computational methods. We have tackled many of these challenges by deploying the sequential Bayesian procedure used by dos Reis et al. (2012) wherein the posterior age estimates derived from a small taxonomic sample with genome-scale data are then deployed as priors for a subsequent analysis with many species and a much-reduced nucleotide sample. This procedure reduces the computational cost of a typical combined data analysis. It also helps to alleviate the concerns with the "missing data" problem as, in our approach, the sequence likelihood is only calculated for the species present in the alignment (Yang and Rannala, 2006).

The molecular timeline for primate evolution that emerges from this study can be interpreted with confidence. The dataset is sufficiently large to provide highly precise branch length estimates, and the methods used are robust in accommodating violations of the molecular clock. The comparison of the two calibration strategies reveals their impact on the results by giving different age estimates within the tree, though the
variation in inferred ages is not extreme. Which ages are considered most accurate will depend in large part on the degree of confidence in the fossils and their placement. As an unanticipated result of the study, the difference in age estimates for the deepest nodes of the phylogeny differ markedly when comparing the molecular rate models, with Bayesian model selection supporting the autocorrelated model. As with previous studies over the past several decades, the ancestral primate lineage is hypothesized to have survived the great K-Pg mass extinction event.

**METHODS**

Bayesian estimates of divergence times of Primates were obtained using a supermatrix of molecular data with 372 species and 3.44 million base pairs (Mbp), combined with 17 fossil calibrations. The matrix is the result of merging the 372-species and 61 thousand base-pairs (Kpb) data set of Springer et al. (2012) with a 10-species subset of the genome-scale alignment of dos Reis, et al. (2012). Bayesian analyses were done using the program MCMCTree (Yang 2007). We assessed the robustness of time estimates by varying the clock model (strict clock, independent rates and correlated rates), and by obtaining estimates under two fossil calibration strategies. Note that time estimates were obtained in two steps: In the first step estimates were obtained for the small phylogeny of 10 species with a long alignment (3.38 Mbp). The marginal posterior of times was then used to construct the time prior in the second step for the 372-species phylogeny with a shorter alignment (61 Kbp). This approach is almost the same as analysing the fully concatenated alignment in one step (3.38 Mbp + 0.061 Mbp), but is computationally less demanding. All alignments, tree topology and fossil calibrations are available in Supplementary Material.

*Sequence Alignment and Tree Topology*

**Springer et al. (2012) alignment.**—We retrieved the sequence alignment of Springer, et al. (2012), which is an extended version of the alignment of Perelman, et al. (2011). The alignment has 372 species (367 primates and 5 outgroup species) and 79 gene segments (69 nuclear and 10 mitochondrial). The composite lagomorph sequence (an outgroup) was removed. We added a scandentian species (*Tupaia belangeri*), because its complete genome is available in the alignment of dos Reis, et al. (2012), and because it has the 10
gene segments from the mitochondrial genome available (accession NC_002521). Many of the nuclear gene segments in the alignment of Springer, et al. (2012) were mixtures of introns, exons and UTRs, with out-of-frame indels in some exons. We manually curated the exons, and separated the coding and non-coding segments of the alignment. These adjustments were necessary to facilitate an informed partition-based analysis of the data. Our modified version of Springer's alignment was thus divided into 6 partitions: (1) 1st and 2nd codon positions for mitochondrial genes; (2) 3rd positions for mitochondrial genes; (3) mitochondrial RNA genes; (4) 1st and 2nd codon positions for nuclear genes; (5) 3rd positions for nuclear genes; and (6) non-coding segments of nuclear genes (UTRs and introns). The concatenated alignment has 372 species and is 61,132 base pairs long (Table 2). Our partitioning into codon positions and coding vs. non-coding sequences follows established recommendations (Shapiro et al., 2006; Yang and Rannala, 2006; Nascimento et al., 2017).

dos Reis et al. (2012) alignment.—We retrieved the genome-scale sequence alignment of dos Reis, et al. (2012) of 36 mammal species, from which we extracted the sequences for 9 primates and 1 scandentian. The dos Reis et al. (2012) alignment was prepared using the highly curated mammalian genomes available in Ensembl. Though three additional primate genomes have become available in this database in the time since the original alignment was prepared, it is unlikely that their inclusion would change our results. The nine species represented in our study provide comprehensive phylogenetic representation of all major nodes in the primate tree and represent each of the higher-level clades (Figure 1, inset). The original alignment has 14,632 nuclear, protein-coding genes, from which we removed 43 genes that were already present in the Springer alignment and 1 gene that was extremely long. All columns in the alignment with ambiguous nucleotides were removed, though care was taken not to disrupt the reading frame of the aligned coding sequences. The alignment was divided into two partitions: (1) 1st and 2nd codon positions; and (2) 3rd codon positions. The final alignment has 10 species and is 3,441,106 base pairs long (missing data 0%, Table 2).

Tree topology.—The topology of the 372-species phylogeny was estimated by maximum likelihood (ML) using RAxML v 8.0.19 (Stamatakis 2014) under the GTR+G model (Yang 1994b, a), using seven partitions (Table 2) and 100 bootstrap replicates.
Fossil Calibrations and Time Prior

The two fossil calibration strategies used in this study are summarized in Table 3. They represent two different interpretations of the fossil record to construct calibrations for use in molecular clock dating analyses. Calibration strategy A is novel to this study, and calibration strategy B is based on the primate calibrations of dos Reis, et al. (2012).

Detailed justifications for the novel calibrations are provided in Appendix 1.

Fossil calibration strategy A.—We used the fossil-based prior densities constructed by Wilkinson, et al. (2011) to calibrate the ages of crown Primates and crown Anthropoidea. The prior densities were constructed by modelling the processes of speciation, extinction, fossil preservation, and rates of fossil discovery in Primates. The effects of the K-Pg extinction were accounted for in the model. We calibrated six more node ages by using uniform distribution densities with soft bounds (Yang and Rannala 2006). We set the probability of violating a minimum bound to 1%. Because maximum bounds are based on weak evidence, we set the probability that a maximum bound is violated to 10% or 20%. The crown Haplorrhini node was left with no calibration as the branch separating that clade from crown Primates is very short and we wanted to avoid truncation with the fossil-modelling density on crown Primates. The prior on the age of crown Haplorrhini is instead set using the birth-death process with parameters $\lambda = \mu = 1$ and $\rho = 0$. These parameter values specify a uniform kernel density (Yang and Rannala 1997, equation 7).

Fossil calibration strategy B.—We used the same nine calibrations that dos Reis, et al. (2012) used to calibrate the Primates and Scandentia clades. An additional calibration based on †Tarsius sp. (Beard, et al. 1994) was used for the Haplorrhini node. For nodes with a minimum bound only, modelled using a truncated Cauchy density, the spread parameter was set to $c = 2$ (Inoue, et al. 2010). For maximum bounds the probability that the bound was violated was set to 5%. There are other differences between strategies A and B (Table 1). For example, in A, we considered †Sahelanthropus, dated to 7.25 million years ago (Ma), to be the oldest member of the human-chimpanzee clade (Brunet et al. 2002, 2005) and used it to calibrate the clade accordingly, while in B, dos Reis, et al. (2012) used †Orrorin (5.7 Ma) instead. In A, †Chororapithecus is given an age
of 10 Ma (Geraads, et al. 2002), while in B it is given the younger (perhaps more

We note that the ages of fossils and their relationships to extant groups are often
controversial and we cannot overemphasize the degree to which differences of opinion
among palaeontologists are an important source of uncertainty in the construction of
fossil calibrations, and accordingly, divergence time estimates throughout the
phylogeny.

Calibrating the 372-species phylogeny. —Strategies A and B were used to obtain time
estimates for the 10-species phylogeny using the 3.38 Mbp alignment. Then skew-t
densities were fitted by ML to the marginal posterior ages of each of the 9 internal nodes
in the 10-species phylogeny, and used to calibrate the corresponding nodes in the 372-
species tree. Eight additional fossil calibrations (Table 3) were used to calibrate
additional nodes in the 372-species tree. For nodes without calibrations, the time prior
was constructed using the birth-death process with parameters $\lambda = \mu = 1$ and $\rho = 0$.
Bayesian time estimation then proceeded on the 372-species tree and 61 Kbp alignment
as usual.

Rate Prior

For the 10-species analysis, the rate prior was set as follows: the nuclear substitution
rate at third codon positions in apes is roughly within $10^{-9}$ substitutions per site per
year (s/s/y) (Burgess and Yang, 2008). At first and second codon positions, it is about a
fifth of the third position rate, or $2 \times 10^{-10}$ s/s/y. This gives roughly an overall rate of
about $5 \times 10^{-10}$ s/s/y for the three positions combined. We thus used a diffuse gamma
density $G(2, 40)$ with mean 0.05 and 95% prior credibility interval (CI) 0.00606–0.139
(our time unit is 100 My, thus, this corresponds to $6.06 \times 10^{-12}$ to $1.39 \times 10^{-10}$ s/s/y). The
analysis was conducted under both the auto-correlated rates (AR) and independent
rates (IR) models. Parameter $\sigma^2$ in the AR and IR models was assigned a gamma prior
$G(1, 10)$. Note that the average rate for loci, $\mu_i$, and $\sigma_i^2$ are assigned a gamma-Dirichlet
prior (dos Reis, et al. 2014b).
For the 372-species phylogeny, the rate prior was assigned as follows: the mitochondrial substitution rate at third positions is about 20 times the rate at third positions in nuclear genes or $2 \times 10^{-8}$. Assuming 1st and 2nd codon positions evolve at about a fifth of the third position rate we get roughly $4 \times 10^{-9}$. The prior mean is then approximately $2.5 \times 10^{-9}$ s/s/y, which is the weighted average (by number of sites) of the substitution rates for the nuclear and mitochondrial partitions. We thus used a gamma density $G(2, 8)$ with mean 0.25 and 95% CI 0.0302–0.696. For $\sigma^2$ we used $G(1, 10)$.

**MCMC and Bayesian Selection of Clock Model**

MCMC analyses were carried out with the program MCMCTree (Yang 2007), using the approximate likelihood method (dos Reis and Yang 2011). Convergence of the MCMC to the posterior distribution was assessed by running the analyses multiple times. MCMCtree runs were carried out without sequence data to calculate the joint prior of node ages. Results from all analyses were summarised as posterior means and 95% CIs.

We have implemented marginal likelihood calculation by thermodynamic integration (path sampling) in the program MCMCTree. This allows us to calculate Bayes factors (BF) and posterior model probabilities to select for a clock model in the analysis. Details of our implementation are given in Appendix 2. Extensive discussions on marginal likelihood estimation by thermodynamic integration and stepping-stones (a related method) in the phylogenetic context are given in Lartillot and Philippe (2006), Lepage, et al. (2007), Xie, et al. (2011) and Baele et al. (2012a, 2012b). A detailed simulation study is given in Ho et al. (2015).

Thermodynamic integration is computationally intensive as we must sample from the power posterior $f(\Theta) f(D|\Theta)^p$, in a sampling path from the prior ($\beta = 0$) to the posterior ($\beta = 1$). Because the approximation to the likelihood is not good when samples are taken far away from the maximum likelihood estimate (dos Reis and Yang 2011), as it happens when $\beta$ is small, the approximation cannot be used in the calculation of the power posterior. Thus, we use exact likelihood calculation on a smaller dataset of nine primate species (inset of Fig. 1), for the six partitions of the Springer alignment (Table 2) to perform the Bayesian selection of clock model. We use 64 $\beta$-points to construct the
sampling path from the prior to the posterior and calculate the marginal likelihoods for
the strict clock, and the AR and IR models.

**Effect of genome-scale data**

In a conventional statistical inference problem, the variance of an estimate decreases in
proportion to $1/n$, with $n$ to be the sample size. Thus, as the sample size approaches
infinity, the variance of an estimate approaches zero and the estimate converges to the
true value. In divergence time estimation, which is an unconventional estimation
problem, the non-identifiability of times and rates means that the uncertainty in the
posterior of times does not converge to zero as the amount of molecular data (the
sample size) approaches infinity, but rather converge to a limiting value imposed by the
uncertainties in the fossil calibrations (Yang and Rannala, 2006; Rannala and Yang,
2007). For infinitely long alignments, an infinite-sites plot (a plot of uncertainty in the
time posterior, measured as the width of the CI, i.e., the difference between the 2.5% and
97.5% limits vs. the mean posterior of time) would converge onto a straight line. This
line represents the amount of uncertainty in time estimates for every 1 million years
(My) of divergence that is due solely to uncertainties in the fossil calibrations. We
calculate the infinite-sites plots for time estimates on the 372-species phylogeny to
study the effect of genome-scale data on the uncertainty of species-level time estimates.

**RESULTS**

* A Timeline of Primate Evolution

Figure 1 illustrates time estimates under the AR model and calibration strategy A (Figs.
2 - 4 show detailed timetrees for the major clades). Under calibration strategy A and the
AR model, we find that crown Primates originated 79.2–70.0 million years ago (Ma),
before the K-Pg event at 66 Ma. However, the diversification of the main clades occurred
much later. Crown Anthropoidea originated 48.3–41.8 Ma, with its two main crown
groups, Catarrhini (Old World monkeys and apes) and Platyrhini (New World
monkeys) originating at 35.1–30.4 Ma and 27.5–23.6 Ma respectively. Crown tarsiers
originated 33.5–15.5 Ma. Crown Strepsirrhini date back to 66.8–58.8 Ma, with its two
main crown groups, Lemuriformes and Lorisiformes, dating back to 61.6–52.7 Ma and
40.9–34.1 Ma respectively.
Calibration strategy B under the AR model gives similar node age estimates for the younger nodes in the tree (i.e. the 95% CI of node age overlap, Fig. 5a). However, for the older nodes in the phylogeny (and in particular for Euarchonta, Primatomorpha, Primates, Haplorrhini, Lemuriformes and Lemuriformes minus aye-aye), strategy A produced older estimates (Figure 5a). Under strategy B a pre-K-Pg origin of crown Primates is also favoured, although the posterior distribution of the age of crown Primates straddles the K-Pg boundary (71.4–63.9 Ma). The posterior probability for a pre-K-Pg origin of crown Primates is 80.0% under strategy B and 100% under strategy A. Posterior time estimates for all nodes under both strategies are given in the Supplementary Material spreadsheet.

Note that the two calibration strategies are in many cases based on the same fossils (Table 3), and the intervals defined by the fossil bounds overlap extensively between the two strategies. However, the seemingly small differences between the two strategies lead to noticeable differences in the posterior time estimates (Fig. 5a). In general, minimum bound constraints are older in strategy A than in strategy B (Table 3), and thus this may be the cause of the older time estimates in A vs. B.

**Time prior and effect of truncation and outgroups.**—User-specified calibration densities usually do not satisfy the constraint that descendant nodes must be younger than their ancestors, thus the dating methodology must ‘truncate’ the calibration densities to satisfy the constraint to construct the time prior (Rannala, 2016). The result is that user-specified calibration densities and marginal priors of node ages may look substantially different. Supplementary Figure 1 illustrates the effect of truncation on prior densities for strategies A and B. For example, in strategy B, the calibration densities on Euarchonta (the root of the phylogeny) and on Primates interact (the primate node has a Cauchy calibration density with a heavy tail), and consequently the prior density on the age of Euarchonta is pushed back (Sup. Fig. 1). The result is that the marginal prior age of Euarchonta ranges from 136–78 Ma (Sup. Fig. 1) instead of 130–61.5 Ma as in the calibration density (Table 3), while the upper age for the Primate prior is too old (127 Ma). In contrast, under strategy A, the calibration density on Primates has a much lighter tail, and thus the truncation effect with the Euarchonta node is
minimal. The result is that the marginal time prior and the corresponding calibration densities for the Primates and Euarchonta nodes are very similar (Sup. Fig. 1). Similarly, under strategy B, the priors for two other nodes (Anthropoidea and Human-Gorilla) that use the heavy-tailed Cauchy calibrations have upper 95% limits that also appear unreasonably old (86.3 Ma and 25.0 Ma respectively). In general, calibration strategy A, which avoids using the long-tailed Cauchy calibrations, has calibration densities that are much closer to the resulting marginal priors, and thus strategy A results in a time prior which is much closer to the fossil information as interpreted by the palaeontologist.

A Set of Calibrations for Mitogenomic Phylogenetic Analysis. Mitogenomic markers are widely used to construct phylogenies of closely related primate species with examples seen in phylogeographic studies of diversification of primates in the Amazon (Nascimento, et al. 2014) and in the timing of human diversification (Rieux, et al. 2014). The posterior distributions obtained here for the 10-species genomic data are useful calibrations for mitogenomic studies. Note that these cannot be used if the molecular alignment contains nuclear data as the calibrations already contain the information from nuclear genomes. The list of skew-t calibrations is provided in Table 4, together with approximations based on the gamma distribution, which can be used in software that does not implement skew-t calibrations (such as BEAST or MrBayes).

Effect of the Clock Model

The clock model has a strong impact on posterior time estimates, particularly for the most ancient nodes in the phylogeny. Under the IR model, the ages of Euarchonta, Primatomorpha, Primates, Haplorrhini and Strepsirrhini are substantially older than those estimated under the AR model (Figure 5b). Posterior means and 95% CIs for locus (partition) rates obtained under both clock models are given in the Supplementary Material spreadsheet.

Results of Bayesian model selection of clock model using thermodynamic integration are shown in Table 5. The AR model has the highest marginal likelihood in 5 out of the 6 partitions analysed, with the posterior model probability > 90% in two partitions, and 79%, 66%, 53% and 29% in the other four. When the six partitions are analysed in a
multi-partition dataset, the posterior probability is virtually 100% in favour of the AR model. We note that ideally, the marginal likelihood calculations should have been carried out on the complete dataset, but unfortunately, this is so computationally expensive that it cannot be done in a feasible amount of time. Thus, further work is necessary to confirm whether the preference for the AR model will remain in analysis of a more taxonomically dense primate phylogeny.

Effect of genome-scale data

Figure 6 shows the infinite-sites plot for the primate data analysed here. For calibration strategy A, the eight primate nodes shared between the 10-species and 372-species trees (i.e. the nodes constrained by the large genome-scale alignment, table 2) fall in an almost perfectly straight line ($R = 0.992$, Fig. 6A). This indicates that for these nodes, uncertainty in the time estimates is dominated by uncertainties in the fossil calibrations rather than by uncertainties in the molecular data. For strategy A, a regression line through the origin fitted to the eight data points is $w = 0.128t$, meaning that for every 1 My of divergence, 0.128 My are added to the CI width (Fig. 6A). On the other hand, when considering all 371 nodes in the tree, the relationship between CI width and mean times is far from linear, and the level of uncertainty is much higher. In this case, 0.277 My are added to the CI width for every 1 My of divergence. The trend is similar under calibration strategy B (Fig. 6B), albeit in this case there is in general more uncertainty in time estimates (i.e. the slope of the regression lines is larger). This appears due to strategy B being more conservative than strategy A, that is, some of the calibration densities used in B are substantially wider, encompassing larger time ranges (Sup. Fig. 1).

DISCUSSION

A Phylogenomic View of Primate Divergences

A primary aim of this work was to study the effect of genome-scale data on divergence time estimates on a species-level phylogeny. Given the wide availability of whole genome data for a core-set of species, it is important to know whether the use of these data for a subsample of lineages will be enough to reduce time uncertainties in a species-level phylogeny to the theoretical limit. The results of figure 6 clearly indicate
this is not the case. Although for the core ancestral nodes in the primate phylogeny, the
genome-scale alignments do constrain uncertainty in time estimates close to their
theoretical limit (so it is highly unlikely that adding additional molecular data for these
species will improve time estimates appreciably), for species without genome-scale
data, there are still substantial uncertainties left for family-level and genus-level
diversions in the tree. For some nodes, the CI-width is almost as large as the node age
(for example, for Tarsiidae, the node age is 23.8 Ma with CI-with 18 My, which is 76% of
the node age, see Supplementary Material spreadsheet). Thus, much work is still
needed in order to improve time estimates for the hundreds of more recent diversences
in the tree. Furthermore, application of morphological-based models for dating
(Ronquist et al. 2012) and the fossilised birth-death process (Heath et al. 2014) also
offer exciting prospects and challenges in obtaining time estimates for the species-level
diversions (O’Reilly et al. 2015, dos Reis et al. 2016). Improving these estimates will
be important in studies of primate diversification rates and to correlate primate
diversification events with major geological events in the history of the Earth (such as
glaciations, continental drift, the closure the Panama isthmus, etc.).

**Sequential Bayesian Analysis versus Secondary Calibrations**

In this work we used the posterior of times obtained under a small dataset as the prior
of times in a second analysis under a large dataset. This approach is justified as long as
the datasets are independent under the likelihood model and as long as the datasets do
not overlap (that is, they share no genes). The use of the posterior in an analysis as the
prior for the next is a well-known feature of Bayesian inference (Gelman et al. 2013).
Consider data that can be split into two subsets, $D = (D_1, D_2)$, which are independent
under the likelihood model. The posterior distribution for parameter $\theta$ is

$$f(\theta \mid D) \propto f(\theta) f(D_1 \mid \theta) f(D_2 \mid \theta) \propto f(\theta \mid D_1) f(D_2 \mid \theta),$$

where $f(\theta \mid D_1) \propto f(\theta) f(D_1 \mid \theta)$ is the posterior distribution of $\theta$ when only $D_1$ are
analysed. It is apparent that using $f(\theta \mid D_1)$ as the prior when analysing $D_2$ leads to the
posterior for the joint data $D$. In other words, performing the analysis in one step (joint
analysis of $D_1$ and $D_2$) or in two steps (posterior under $D_1$ as prior under $D_2$) results in
the same posterior distribution.
The approach we used here to analyse the primate data is justified because the likelihood model assumes that the sequence partitions are non-overlapping and independent. However, our approach is approximate. In multi-parameter models, the posterior is a multidimensional distribution that may have a complex correlation structure. Here we ignored the correlation structure of the nine times estimated using the genomic data, and approximated the corresponding high-dimensional time posterior as the product of the marginal densities of the times, with a truncation applied to ensure that descendants are younger than ancestors. Note that joint analysis of all the partitions would have been preferable, but it is computationally prohibitive.

This Bayesian sequential analysis is different from the use of secondary calibrations in some dating studies (Graur and Martin, 2004), where the secondary calibrations were used as point calibrations (with the uncertainties on node age estimates ignored), and where in many cases the data analysed under the secondary calibration was the same as the data analysed to obtain the calibration in the first place (Graur and Martin, 2004).

**Clock Model**

An interesting result from our study is the finding that the AR model fits the primate data better than the IR relaxed-clock model. In the context of previous studies, Lepage, et al. (2007) found, using Bayes factors and no fossil calibrations, that two AR relaxed clocks (CIR and log-normal) fitted real data (eukaryotes, mammals and vertebrates) better than IR models. More recently, Lartillot, et al. (2016) introduced a mixed relaxed clock that has auto-correlated- and independent-rates components. In their analysis, the mixed clock appeared to provide a better description of rate evolution in the mammal phylogeny, however, they did not assess clock model fit with Bayes Factors. Linder et al. (2011) found, also by using Bayes factors, that IR models better fit an angiosperm phylogeny better than AR models. Additionally, they found that, when analysed without fossil calibrations, the AR model fit an ape phylogeny better than the IR model. However, when analysed with fossil calibrations, the IR model fit the ape data better.
In the AR model the variance of the log-rate for branches is proportional to the time of divergence, so that the variance is expected to be close to zero for closely related species. In other words, the AR model allows for “local clocks” for closely related species, while allowing the rate to drift substantially across distantly related clades. This model is, from a biological point of view, quite appealing intuitively, and it also fits anecdotal evidence where the strict clock cannot be statistically rejected among very closely related species, for example, among the apes (dos Reis, et al. 2016, box 2). In contrast, the IR model assumes that the variance of the branch rates is the same for different time scales. This would appear biologically unrealistic. Arguments have been put forward in favour and against both of the two types of relaxed-clock models examined by our study (Thorne et al. 1998, Drummond et al. 2006, Ho 2009), and clearly further research is still needed to understand the merits of each clock model and how they relate to evolutionary patterns in different genomic regions (Ho 2014). This will be a challenging task given how difficult it has been to distinguish between the two models in simulation studies (Heath et al. 2012, Ho et al. 2015, Lepage et al. 2007).

**Five Decades of Primate Molecular Timetrees**

Prior to the study reported here, the estimated age of the living primate clade has spanned a 30 My differential, ranging from as young as 55 Ma (Koop et al., 1989) to as old as 87 Ma (Perelman et al, 2011). The new millennium has been a particularly active time for primate divergence time analysis. Beginning in the early 2000's, published studies have employed a diverse assortment of datasets applied to the problem (e.g., nuclear, mitochondrial, and their combination), as well as a range of statistical methods and calibration densities. Despite this array of data and methods, all of these studies—with only one notable outlier (Chatterjee, 2009)—have consistently indicated that the crown Primates clade originated prior to the K-Pg event (see also Steiper and Seiffert, 2012). Given the continued dearth of fossil data to support this hypothesis, however, the result continues to be viewed with scepticism by the paleoanthropological community (Bloch et al., 2007; Silcox, 2008; O’Leary et al., 2013; but see Martin et al., 2007).

As described at length above, the current study gives added weight to the conclusion that primates are an ancient clade of placental mammals, arising just prior to or millions
of years before the K-Pg. And even though lineage diversification within the major
subclades is hypothesized not to have occurred until after the commencement of the
Paleogene, the separation of tarsiers from other haplorrhines, and the divergence of
haplorrhines and strepsirrhines, consistently appear to proceed or nearly coincide with
the K-Pg. Given that this event was unequivocally one of the most disruptive and
destructive geological episodes in Earth history, the temporal coincidence speaks both
to the ecological flexibility and to the evolutionary opportunism of the earliest primates.
Although now extinct in North America and in Europe, the primate fossil record shows
that the clade was once nearly pan-global, even potentially including Antarctica. Thus,
when viewed in the context of divergence date estimates, all of which fall within a
temporal window when, as now, continental and island residences would have already
been sundered by significant oceanic barriers (most notably, the separation of South
America from Africa by the Atlantic Ocean), we must conclude that early primates would
have been able dispersers. In fact, the ability to cross barriers, both terrestrial and
aquatic, and to successfully colonize new land masses, are distinct hallmarks of the
primate radiation (Gingerich, 1981; Yoder and Nowak, 2006; de Queiroz, 2005, 2014;
Seiffert, 2012; Beard, 2016; Bloch et al., 2016).

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository:
http://dx.doi.org/10.5061/dryad.c020q. An R package, mcmc3r, that helps the user
with marginal likelihood estimation with MCMCTree is available from
https://github.com/dosreislab/mcmc3r.

FUNDING
This work was supported by grant BB/J009709/1 from the Biotechnology and
Biosciences Research Council (UK). Part of this work was carried out while MdR was
visiting the National Evolutionary Synthesis Center (NESCent, National Science
Foundation #EF-0905606) in summer 2013. JBM was supported by a joint scholarship
from University College London and the Government of Mexico’s CONACYT. ADY was
supported by a grant from the Burroughs Wellcome Fund and by DEB-1354610 from the
National Science Foundation.
DEDICATION

This paper is dedicated to the memory of our co-author Gregg F. Gunnell.

APPENDIX 1

Justifications for dates assigned to 17 fossil calibrations in this study (Table 3) are given below: 8 calibrations for strategy A (SA), 1 calibration for strategy B (SB), and 8 calibrations shared among both strategies (SAB). The justifications for the remaining calibrations in Table 3 are given in dos Reis et al. (2012; see also Benton et al. 2009). In some cases, the dates used are not exactly those published in cited references. In these cases, the dates utilized reflect published as well as unpublished information or adjustments deemed necessary given the uncertainty of some dates. In any case the discrepancies are always small and unimportant considering the breadth of the fossil calibrations. Note that specifying maximum bounds is a difficult task, in particular because absence of fossil evidence is not evidence that a clade did not exist in a point in time (Ho and Philips, 2009). We use stem fossils as benchmark points onto which to construct diffuse maximum bounds (i.e. with a large probability of violation, $p_U$) on some node ages.

Hominini | Homo-Pan | 7.5 Ma – 10 Ma | SA

The minimum age for the divergence of hominins is placed at 7.5 Ma and is based on the appearance of †Sahelanthropus at 7.2 Ma (Brunet, et al. 2002, 2005, Lebatard, et al. 2008). There is some controversy as to the proper taxonomic position of Sahelanthropus (Wolpoff, et al. 2006, MacLatchy, et al. 2010) but we regard it as the oldest record of a plausible crown hominin. Sahelanthropus comes from the Anthracotheriid Unit of an unnamed formation in the Mega-Chad Basin in Chad (Brunet, et al. 2005). The associated mammalian fauna is very similar to that found from the Nawata Formation, Lothagam, Kenya which may be as old as 7.4 Ma (MacDougall 2003). The divergence of the hominin lineage seems unlikely to have occurred before the appearance of the potential gorillin Chororapithecus at 10 Ma (Suwa, et al. 2007, Harrison 2010a).

Homininae | Gorilla-Homo | 10 Ma – 13.2 Ma | SA
The minimum age for the divergence of crown hominines is placed at 10 Ma based on the appearance of the potential gorilline *Chororapithecus*. Like *Sahelanthropus*, the taxonomic status of *Chororapithecus* is not without controversy (Suwa et al. 2007, Harrison 2010a). Harrison (2010a) regards *Chororapithecus* as best interpreted as a stem hominin or even a stem hominid – we feel that the features that do support a relationship with gorillas are well enough established to use the date of appearance of *Chororapithecus* as a minimum divergence date for hominines. *Chororapithecus* comes from the late Miocene Beticha section of the Chorora Formation in Ethiopia and is dated at 10-10.5 Ma (Geraads, et al. 2002). We use a maximum divergence date of 13.2 (Raza, et al. 1983) for *Sivapithecus* but it is now evident that this date might be slightly too old. In light of this the divergence of the hominine lineage is unlikely to have taken place before the earliest appearance of the probable crown pongine *Sivapithecus* at 12.5 Ma (Begun 2010, Begun, et al. 2012).

**Hominidae | Pongo-Homo | 11.2 Ma – 28 Ma | SA**

The minimum age for the divergence of crown hominids is placed at 11.2 Ma based on the earliest appearance of the crown pongine *Sivapithecus* (Kappelman, et al. 1991, Begun 2010, Begun, et al. 2012). *Sivapithecus* is known from Siwalik group rocks (Chinji, Nagri and Dhok Pathan formations) in Indo-Pakistan that range in age from 14 Ma to 5.5 Ma (Badgley and Behrensmeyer 1995) with *Sivapithecus* restricted to a range of 12.5 Ma to 7.4 Ma (Flynn, et al. 1995). The divergence of crown hominids is unlikely to have occurred before the first appearance of *Kamoyapithecus* at 25 Ma (Seiffert 2010). We use 28 Ma as a slightly more conservative maximum.

**Catarrhini | Homo-Macaca | 25 Ma – 33.7 Ma | SA**

The presence of the crown hominoid *Kamoyapithecus* (Zalmout, et al. 2010) indicates that the minimum divergence time for crown Catarrhini is 25 Ma (Seiffert 2010). *Kamoyapithecus* is only known from the Erageliet beds, Kalakol Basalts locality of Lothidok in Kenya (Madden 1980, Leakey, et al. 1995, Rasmussen and Gutierrez 2009). A soft maximum of 33.7 Ma on the age of crown catarrhines is given due to the absence of hominoids before 33.7 Ma.

**Anthropoidea | Catarrhini-Platyrrhini | 41 Ma – 62.1 Ma | SA**
Calibration density constructed from fossil modeling. The effects of the K-Pg extinction are included in the model. See Wilkinson, et al. (2011) for details.

Haplorrhini | crown Tarsius | 45 Ma | SB

The presence of the crown tarsid Xanthorhysis in Shanxi Province, China (Beard 1998) and apparently of the genus Tarsius in fissure fills at Shanghuang in Jiangsu Province, China (Beard, et al. 1994), both dating to the late middle Eocene (40-45 Ma), circumscribe the minimum divergence time for crown Haplorrhini at 45 Ma.

Strepsirrhini | Lorisiformes-Lemuriformes | 37 Ma – 58 Ma | SA

The minimum age for the divergence of crown Strepsirrhini is placed at 37 Ma based on the first appearance of the crown lorisiform Saharagalago (Seiffert, et al. 2003). Saharagalago is only known from Fayum Quarry BQ-2 in the Birket Qarun Formation, Egypt. The divergence of crown strepsirrhines is unlikely to have occurred before the first appearance of the basal primate Altiatlasius from Ouarzazate in Morocco which is considered to represent the late Paleocene (Thanetian) and dating to around 58 Ma (Gheerbrant, et al. 1993, Gheerbrant, et al. 1998, Seiffert 2010).

Primates | Haplorrhini-Strepsirrhini | 57.6 Ma – 88.6 Ma | SA

Calibration density constructed from fossil modeling. Includes the effects of the K-Pg extinction in the model. See Wilkinson, et al. (2011) for details.

Euarchonta | Scandentia-Primates | 65 Ma – 130 Ma | SA

The minimum age for the divergence of crown Euarchonta is placed at 65 Ma based on the first appearance of the crown euarchontan Purgatorius (Bloch, et al. 2007). Purgatorius is known from the early Paleocene (Puercan) Tullock and Bear Formations in Montana (Clemens 1974, Buckley 1997, Clemens 2004, Chester, et al. 2015) and from the earliest Paleocene Ravenscrag Formation in Saskatchewan (Fox and Scott 2011). The divergence of Euarchonta is unlikely to have been before the appearance of placental mammals by at least 130 Ma (Luo 2007, but see Luo, et al. 2011 for a potential 130 My old eutherian).

Lorisidae | Nycticebus-Perodicticus | 14 Ma – 37 Ma | SAB
The minimum age for the divergence of crown Lorisidae is placed at 14 Ma based on an
undescribed genus and species from Fort Ternan in Kenya cited by Walker (1978) and
Harrison (2010b). The minimum age could possibly be as old as 19 Ma if *Mioeuoticus*
(Leakey in Bishop 1962, Walker 1978) represents a crown lorisid (Harrison 2010b).
Fossil lorisids are known from the early to middle Miocene in Africa (Phillips and
Walker 2000, 2002, Harrison 2010b) and from the late Miocene of Pakistan (Jacobs
1981). The divergence of lorisoids is unlikely to have occurred before the first
appearance of the potential stem lorisid *Karanisia* at BQ-2 in Egypt (Seiffert, et al. 2003).

**Galagidae | Galago-Euoticus | 15 Ma – 37 Ma | SAB**

The minimum age for the divergence of crown Galagidae is placed at 15 Ma based on an
undescribed genus and species from Maboko Island in Kenya cited by McCrossin (1999)
and Harrison (2010b). The minimum age could possibly be as old as 19 Ma if either
*Progalago* or *Komba* represent a crown galagid (MacInnes 1943, Simpson 1967,
Harrison 2010b). Fossil galagids are known from the early Miocene through early
Pleistocene in Africa (Phillips and Walker 2002, Harrison 2010b). The divergence of
galagids is unlikely to have occurred before the first appearance of the potential stem
lorisid *Karanisia* at BQ-2 in Egypt (Seiffert, et al. 2003).

**Lorisiformes | Galago-Perodicticus | 18 Ma – 38 Ma | SAB**

The minimum age for the divergence of crown Lorisiformes is placed at 18 Ma based on
the appearance of the potential crown lorisoid *Mioeuoticus* in the early Miocene of East
Africa (Harrison 2010b). The divergence of lorisiforms is unlikely to have occurred
before the first appearances of *Karanisia* at BQ-2 in Egypt (Seiffert, et al. 2003). We use
38 Ma as a conservative soft maximum.

**Platyrrhini | Pitheciiidae-Callitrichidae | 15.7 Ma – 33 Ma | SAB**

The minimum age for the divergence of crown Platyrrhini is based on the first
occurrence of the crown pitheciine *Proteropithecia* dated at 15.7 Ma (Kay, et al. 1998,
Fleagle and Tejedor 2002). The minimum age could be as much 18 Ma if either (or both)
*Soriacebus* or *Carlocebus* represent crown pitheciins (Fleagle, et al. 1987, Fleagle 1990,
Bown and Fleagle 1993, Fleagle, et al. 1995, Rosenberger 2011). All of these taxa are
known from the early and middle Miocene of Argentina (Fleagle and Tejedor 2002). The
divergence of platyrrhines is unlikely to have occurred before the appearance of the
crown catarrhine *Catopithecus* (33 Ma, Fayum, Egypt) although a recently published
report has claimed a 36 Ma date for a stem platyrrhine from Peru (Bond, et al. 2015). It
remains unclear how this older date was derived, however, and requires further
substantiation.

**Atelidae | Ateles-Alouatta | 12.8 Ma – 18 Ma | SAB**

The minimum age for the divergence of crown Atelidae (as recognized by Rosenberger, 2011; subfamily Atelinae of others) is based on the first appearance of the crown atelid *Stirtonia* (Hershkovitz 1970, Rosenberger 2011) at 12.8 Ma. Fossil atelids are known from the middle Miocene of Colombia and the Quaternary of Brazil and the Greater Antilles (MacPhee and Horovitz 2002). The divergence of atelids is unlikely to have occurred before the first appearance of the potential stem or crown atelid *Soriacebus* at 18 Ma (Bown and Fleagle 1993, Fleagle, et al. 1995).

**Cebidae | Cebus-Saimiri | 12.8 Ma – 18 Ma | SAB**

The minimum age for the divergence of crown Cebidae is based on the first appearance of the crown cebid *Neosaimiri* (Stirton 1951, Hartwig and Meldrum 2002) dated at 12.8 Ma from La Venta, Colombia. There are several older potential crown cebids including *Dolichocebus* and *Tremacebus* from Argentina and *Chilecebus* from Chile, all dated to around 20 Ma but it remains unclear how these taxa relate to the crown group. Recently, an additional potential crown cebid has been described from Panama (Bloch et al. 2016) dated at 20.9 Ma, which if substantiated would push the potential maximum bound to at least 21 Ma. Here we have accepted the notion that the divergence of cebids is unlikely to have occurred before the first appearance of the potential stem or crown atelid *Soriacebus* at 18 Ma (Bown and Fleagle 1993, Fleagle, et al. 1995) but acknowledge that this date could be as old as 21-22 Ma.

**Cercopithecidae | Papionini-Cercopithecini | 5 Ma – 23 Ma | SAB**

The minimum age for the divergence of crown Cercopithecidae is based on the first appearance of *Parapapio* in the late Miocene (5 Ma) at Lothagam, Kenya. It is potentially possible that some specimens of *Parapapio* from Lothagam could be as old as 7.4 Ma (Jablonski and Frost 2010). The divergence of cercopithecids is unlikely to have
occurred before the first appearance of the stem cercopithecoid *Prohylobates*. The oldest documented *Prohylobates* specimens are from Wadi Moghra in the Qatarr Depression, Egypt dated to approximately 19.5 Ma. However, we view *Kamoyapithecus*, dated at 25 Ma as a crown hominoid, which implies that at least stem cercopithecoids were in existence at that time. Given the controversial nature of our views on *Kamoyapithecus*, we have used a date of 23 Ma as the likely maximum divergence time for crown cercopithecoids.

**Colobinae | Colobini-Presbytini | 9.8 Ma – 23 Ma | SAB**

The minimum age for the divergence of crown Colobinae is based on the first appearance of *Microcolobus* in the middle Miocene (9.8 Ma) at Ngeringerowa, Kenya (Benefit and Pickford 1986, Jablonski and Frost 2010). The divergence of colobines is unlikely to have occurred before the first appearance of the stem cercopithecoid *Prohylobates*. The oldest documented *Prohylobates* specimens are from Wadi Moghra in the Qatarr Depression, Egypt dated to approximately 19.5 Ma. However, we view *Kamoyapithecus*, dated at 25 Ma as a crown hominoid, which implies that at least stem cercopithecoids were in existence at that time. Given the controversial nature of our views on *Kamoyapithecus*, we have used a date of 23 Ma as the likely maximum divergence time for crown colobines.

**APPENDIX 2**

Here we briefly describe our new implementation of Bayes factor calculation in MCMCTree. Our approach is the thermodynamic integration-Gaussian quadrature method implemented recently in the program BPP. The mathematical details are given in Rannala and Yang (2017).

In this paper, we compare different rate models, $m_i$, which differ only in the density of the branch rates while the prior on divergence times and the sequence likelihood are the same between the models. The posterior distribution of times ($t$) and rates ($r$) given the sequence data $D$, and given a clock model $m_i$ is thus

$$f(t, r | D, m_i) = \frac{1}{z_i} f(t) f(r | t, m_i) f(D | t, r),$$
where \[ z_i = \int f(t) f(r \mid t, m_i) f(D \mid t, r) \, dt \, dr \]
is the marginal likelihood of the data for model \( m_i \). Let \( m_r \) be the model with highest marginal likelihood, and let BF\( r \) = \( z_i / z_r \) be the Bayes factor of model \( i \) over model \( r \). Then the posterior probability of model \( i \) is

\[
Pr(m_i \mid D) = \frac{z_i \Pr(m_i)}{\sum_j z_j \Pr(m_j)} = \frac{z_i \Pr(m_i)}{\sum_j z_j \Pr(m_j)} = \sum_j \frac{BF_r \Pr(m_i)}{BF_r \Pr(m_j)},
\]

where the sum is over all models being tested, and \( \Pr(m_i) \) is the prior model probability.

We calculate \( z_i \) by sampling from the power posterior

\[
f_\beta(t, r \mid D, m_i) \propto f(t) f(r \mid t, m_i) f(D \mid t, r)^\beta, \quad 0 \leq \beta \leq 1
\]

for a given \( \beta \) value. We choose \( K \beta_j \) values to integrate between 0 and 1 according to the Gauss-Legendre quadrature rule. The estimate of \( \log z_i \) is then given by the quadrature formula

\[
\log z_i \approx \frac{1}{2} \sum_{j=1}^K w_j \bar{\ell}_{\beta_j},
\]

where \( w_j \) are the Gauss-Legendre quadrature weights, and \( \bar{\ell}_{\beta_j} \) is the average of log-likelihood values, \( \ell_{\beta_j} \), sampled from the power posterior with \( \beta_j \). The standard error of the estimate is given by

\[
\text{S.E.} = \frac{1}{2} \sqrt{\sum_{j=1}^K w_j^2 \text{Var}(\ell_{\beta_j}) / \text{ESS}},
\]

where the variance is calculated over the MCMC sample of \( \ell_{\beta_j} \), and \( \text{ESS} \) is the effective-sample size of \( \ell_{\beta_j} \).
REFERENCES


Table 1. Overview of estimates of divergence times in Primates (in millions of years ago) for selected studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Data/Analysis</th>
<th>Primates</th>
<th>Haplorrhini</th>
<th>Anthropoidea</th>
<th>Platyrrhini</th>
<th>Catarrhini</th>
<th>Homininae</th>
<th>Strepsirrhini</th>
<th>Lorisiformes</th>
<th>Lemuriformes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarich and Wilson (1967)</td>
<td>Immunological distance/strict clock</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hasegawa, et al. (1985)</td>
<td>896 bp mtDNA/strict clock</td>
<td>55</td>
<td>40</td>
<td>25</td>
<td></td>
<td>14.7</td>
<td>8.1</td>
<td>41.8</td>
<td>22.1</td>
<td>39.6</td>
</tr>
<tr>
<td>Koop, et al. (1989)</td>
<td>Immunological distance/strict clock</td>
<td>55</td>
<td>40</td>
<td>25</td>
<td></td>
<td>14.7</td>
<td>8.1</td>
<td>41.8</td>
<td>22.1</td>
<td>39.6</td>
</tr>
<tr>
<td>Purvis (1995)</td>
<td>Super Tree Analysis</td>
<td>57.2</td>
<td>39.9</td>
<td>47.6</td>
<td>6.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumar and Hedges (1998)</td>
<td>658 nuclear genes</td>
<td>56</td>
<td>58</td>
<td>40</td>
<td></td>
<td>7</td>
<td>50</td>
<td>23</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Goodman, et al. (1998)</td>
<td>60-80 Kbp globin gene region/strict clock</td>
<td>63</td>
<td>58</td>
<td>40</td>
<td></td>
<td>7</td>
<td>50</td>
<td>23</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Yang and Yoder (2003)</td>
<td>2404 bp mtDNA/local clocks</td>
<td>57.6</td>
<td></td>
<td></td>
<td>7</td>
<td>50</td>
<td>23</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poux and Douzery (2004)</td>
<td>1278 bp nDNA/local clocks</td>
<td>77.2</td>
<td>56.7, 58.4</td>
<td>43.6</td>
<td></td>
<td></td>
<td></td>
<td>45.4, 46.7</td>
<td>13.8, 14.2</td>
<td>39.6, 40.7</td>
</tr>
<tr>
<td>Eizirik, et al. (2004)</td>
<td>8,182 bp nDNA/relaxed clock</td>
<td>77.5</td>
<td>42.9</td>
<td>20.8</td>
<td>30.5</td>
<td>8.6</td>
<td></td>
<td>57.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steiper and Young (2006)</td>
<td>59.8 Kbp of genomic data/relaxed clock w/ autocorrelated rates</td>
<td>77.5</td>
<td>42.9</td>
<td>20.8</td>
<td>30.5</td>
<td>8.6</td>
<td></td>
<td>57.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bininda-Emonds, et al. (2007)</td>
<td>51,089 bp mtDNA &amp; nDNA/ad hoc relaxed clock</td>
<td>87.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chatterjee, et al. (2009)</td>
<td>6,138 bp mtDNA &amp; 2,157 bp nDNA/relaxed clock</td>
<td>63.7</td>
<td>42.8</td>
<td>26.6</td>
<td>23.4</td>
<td>10.7</td>
<td>51.6</td>
<td>37.5</td>
<td>46.2</td>
<td></td>
</tr>
<tr>
<td>Perelman, et al. (2011)</td>
<td>34,927 bp DNA/relaxed clock w/independent rates</td>
<td>87.2</td>
<td>81.3</td>
<td>43.5</td>
<td>24.8</td>
<td>31.6</td>
<td>8.3</td>
<td>68.7</td>
<td>40.3</td>
<td>58.6</td>
</tr>
<tr>
<td>Springer, et al. (2012)</td>
<td>69 nDNA genes; 10 mtDNA genes/relaxed clock w/autocorrelated rates</td>
<td>67.8</td>
<td>61.2</td>
<td>40.6</td>
<td>23.3</td>
<td>25.1</td>
<td>8.0</td>
<td>54.2</td>
<td>34.7</td>
<td>50.0</td>
</tr>
<tr>
<td>dos Reis, et al. (2012)</td>
<td>14,644 genes w/ 20.6 Mbp/relaxed clock w/autocorrelated rates</td>
<td>68.2</td>
<td>65.0</td>
<td>37.4</td>
<td>26.4</td>
<td>10.4</td>
<td>55.1</td>
<td>35.6</td>
<td>49.3</td>
<td></td>
</tr>
<tr>
<td>Finstermeier, et al. (2013)</td>
<td>complete mtDNA genomes/relaxed clock w/autocorrelated rates</td>
<td>66.2</td>
<td>63.0</td>
<td>45.3</td>
<td>22.0</td>
<td>32.0</td>
<td>8.4</td>
<td>56.9</td>
<td>34.5</td>
<td>47.1</td>
</tr>
<tr>
<td>Pozzi, et al. (2014)</td>
<td>complete mtDNA genomes/relaxed clock w/autocorrelated rates</td>
<td>74.1</td>
<td>70.0</td>
<td>46.7</td>
<td>20.9</td>
<td>32.1</td>
<td>10.6</td>
<td>66.3</td>
<td>40.3</td>
<td>59.6</td>
</tr>
<tr>
<td>This Study (Strategy A)</td>
<td>3.4 million bp/relaxed clock w/autocorrelated rates</td>
<td>74.4</td>
<td>70.6</td>
<td>45.0</td>
<td>25.3</td>
<td>32.6</td>
<td>10.5</td>
<td>62.7</td>
<td>37.9</td>
<td>57.2</td>
</tr>
<tr>
<td>This Study (Strategy B)</td>
<td>3.4 million bp/relaxed clock w/independent rates</td>
<td>84.8</td>
<td>78.6</td>
<td>45.9</td>
<td>25.3</td>
<td>28.8</td>
<td>7.6</td>
<td>64.0</td>
<td>38.2</td>
<td>55.3</td>
</tr>
</tbody>
</table>

Notes: All estimates are mean estimates; see original works for confidence/credible intervals; all taxonomic designations signify crown nodes (Primates = the divergence of Strepsirrhini and Haplorrhini; Haplorrhini = the divergence of Tarsiidae from Anthropoidea; Platyrrhini; Catarrhini; Homininae = the divergence of Gorilla from Pan+Homo; Strepsirrhini = the divergence of Lorisiformes and Lemuriformes; Lorisiformes = the divergence of Galagidae and Lorisidae; Lemuriformes = the divergence of Daubentonia (the aye-aye) from other Malagasy lemurs.)
Table 2. Sequence alignment summary

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Partitiona</th>
<th>Sites</th>
<th>Species</th>
<th>Missing datab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Springer et al.</td>
<td>1. mit 1st+2nd</td>
<td>4,816</td>
<td>330</td>
<td>61.4%</td>
</tr>
<tr>
<td></td>
<td>2. mit 3rd</td>
<td>2,408</td>
<td>330</td>
<td>61.4%</td>
</tr>
<tr>
<td></td>
<td>3. mit RNA</td>
<td>2,169</td>
<td>220</td>
<td>45.7%</td>
</tr>
<tr>
<td></td>
<td>4. nuclear 1st+2nd</td>
<td>16,309</td>
<td>239</td>
<td>53.8%</td>
</tr>
<tr>
<td></td>
<td>5. nuclear 3rd</td>
<td>8,156</td>
<td>239</td>
<td>53.8%</td>
</tr>
<tr>
<td></td>
<td>6. nuclear non-coding</td>
<td>27,274</td>
<td>220</td>
<td>46.3%</td>
</tr>
<tr>
<td>Partitions 1–6</td>
<td></td>
<td>61,132</td>
<td>372</td>
<td>51.1% (68.8%)</td>
</tr>
<tr>
<td>dos Reis et al.</td>
<td>7. nuclear 1st+2nd</td>
<td>2,253,316</td>
<td>10</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>8. nuclear 3rd</td>
<td>1,126,658</td>
<td>10</td>
<td>0.0%</td>
</tr>
<tr>
<td>Partitions 7–8</td>
<td></td>
<td>3,379,974</td>
<td>10</td>
<td>0.0% (97.3%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>3,441,106</td>
<td>372</td>
<td>0.78% (96.8%)</td>
</tr>
</tbody>
</table>

a. For topology estimation with RAxML, we used seven partitions: partitions 1 to 3, then 4 and 7 as one partition, 5 and 8 as one partition, and partition 6 divided into two: UTRs and Introns. b. Numbers in brackets are the % missing data for the RAxML analysis. Note that MCMCTree only uses the species present in a partition to calculate the likelihood for the partition. In RAxML missing species in a partition are represented as sequences of only gaps in the partition, and thus the amount of missing data is larger.
Table 3. Fossil calibrations used in this study.

<table>
<thead>
<tr>
<th>Calibration Strategy</th>
<th>Crown Group</th>
<th>Minimum (Ma)</th>
<th>Maximum (Ma)</th>
<th>MCMCTree Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Strategy A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-Chimp</td>
<td>7.5 (†Sahelanthropus)</td>
<td>10 (unlikely before stem gorilla</td>
<td>B(0.075, 0.10, 0.01, 0.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>†Chororapithecus</td>
<td>†Chororapithecus</td>
<td></td>
</tr>
<tr>
<td>Human-Gorilla</td>
<td>10 (†Chororapithecus)</td>
<td>13.2 (unlikely before stem hominin</td>
<td>B(0.10, 0.132, 0.01, 0.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>†Sivapithecus</td>
<td>†Sivapithecus</td>
<td></td>
</tr>
<tr>
<td>Hominidae</td>
<td>11.2 (†Sivapithecus)</td>
<td>28 (unlikely before stem hominoid</td>
<td>B(0.112, 0.28, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>†Kamoyapithecus</td>
<td>†Kamoyapithecus</td>
<td></td>
</tr>
<tr>
<td>Catarrhini</td>
<td>25 (†Kamoyapithecus)</td>
<td>33.7 (absence of hominoids)</td>
<td>B(0.25, 0.337, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthropoidea</td>
<td>41 (K-Pg fossil modeling)</td>
<td>62.1 (K-Pg fossil modeling)</td>
<td>ST(0.4754, 0.0632, 0.98, 22.85)</td>
<td></td>
</tr>
<tr>
<td>Strepsirrhini</td>
<td>37 (†Saharagalago)</td>
<td>58 (unlikely before †Altiatlasius)</td>
<td>B(0.37, 0.58, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Primates</td>
<td>57.6 (K-Pg fossil modeling)</td>
<td>88.6 (K-Pg fossil modeling)</td>
<td>S2N(0.474, 0.65, 0.0365, -3400, 0.6502, 0.1375, 11409)</td>
<td></td>
</tr>
<tr>
<td>Euarchonta</td>
<td>65 (†Purgatorius)</td>
<td>130 (absence of placentals)</td>
<td>G(36, 36.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Strategy B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimp-Human</td>
<td>5.7 (†Orrorin)</td>
<td>10 (absence of hominines)</td>
<td>B(0.057, 0.10, 0.01, 0.05)</td>
<td></td>
</tr>
<tr>
<td>Gorilla-Human</td>
<td>7.25 (†Chororapithecus)</td>
<td>-</td>
<td>L(0.0725, 0.1, 2)</td>
<td></td>
</tr>
<tr>
<td>Hominidae</td>
<td>11.2 (†Sivapithecus)</td>
<td>33.7 (absence of pongines)</td>
<td>B(0.112, 0.337, 0.05, 0.05)</td>
<td></td>
</tr>
<tr>
<td>Catarrhini</td>
<td>23.5 (†Proconsul)</td>
<td>34 (absence of hominoids)</td>
<td>B(0.235, 0.34, 0.01, 0.05)</td>
<td></td>
</tr>
<tr>
<td>Anthropoidea</td>
<td>33.7 (†Catopithecus)</td>
<td>-</td>
<td>L(0.337, 0.1, 2)</td>
<td></td>
</tr>
<tr>
<td>Haplorrhini a</td>
<td>45 (†Tarsius)</td>
<td>-</td>
<td>L(0.45, 0.1, 2)</td>
<td></td>
</tr>
<tr>
<td>Strepsirrhini</td>
<td>33.7 (†Karanisia)</td>
<td>55.6 (absence of strepsirrhines)</td>
<td>B(0.337, 0.556, 0.01, 0.05)</td>
<td></td>
</tr>
<tr>
<td>Primates</td>
<td>55.6 (†Altiatlasius)</td>
<td>-</td>
<td>L(0.556, 0.1, 2)</td>
<td></td>
</tr>
<tr>
<td>Euarchonta</td>
<td>61.5 (carpolestids and plesiadaps)</td>
<td>130 (absence of placentals)</td>
<td>B(0.615, 1.30, 0.01, 0.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shared by both strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorises</td>
<td>14 (Lorisidae gen et sp. nov)</td>
<td>37 (unlikely before †Karanisia)</td>
<td>B(0.14, 0.37, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Galagos</td>
<td>15 (Galagidae gen et sp. nov)</td>
<td>37 (unlikely before †Karanisia)</td>
<td>B(0.15, 0.37, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Lorisiformes</td>
<td>18 (†Mioeouticus)</td>
<td>38 (unlikely before †Karanisia/†Saharagalago)</td>
<td>B(0.18, 0.38, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Platyrrhini</td>
<td>15.7 (stem Pitheciinae)</td>
<td>33 (unlikely before †Catopithecus)</td>
<td>B(0.157, 0.33, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>12.8 (†Stirtonia)</td>
<td>18 (unlikely before †Soriacebus)</td>
<td>B(0.128, 0.18, 0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>12.8 (†Neosaimiri)</td>
<td>18 (unlikely before †Soriacebus)</td>
<td>B(0.128, 0.18, 0.01, 0.10)</td>
<td></td>
</tr>
</tbody>
</table>

37
Cercopithecinae$^c$ 5 (†Parapapio)  23 (unlikely before †Prohylobates/Kamoyapithecus)  B(0.05, 0.23, 0.01, 0.10)

Colobinae$^c$  9.8 (†Microcolobus)  23 (unlikely before †Prohylobates/Kamoyapithecus)  B(0.098, 0.23, 0.01, 0.10)

a. Calibration strategies A and B are applied to the nodes in the phylogeny of 10 species. The shared calibrations are applied to the large tree of 372 species. Ages are in millions of years ago (Ma).

b. B($t_L$, $t_U$, $p_L$, $p_U$) means the node age is calibrated by a uniform distribution bounded between a minimum time $t_L$, and a maximum time $t_U$, with probabilities $p_L$ and $p_U$ that the age is outside the bounds. U($t_L$, $p$, $c$, $p_L$) means the node age is calibrated by a truncated Cauchy distribution with minimum age $t_L$ and parameters $p$ and $c$, with the probability that the age is younger than the minimum bound to be $p_L = 5\%$ (Inoue, et al. 2010). ST($a$, $b$, $c$, $d$) means the node age is calibrated by a skew-t density with parameters $a$, $b$, $c$ and $d$ (Wilkinson, et al. 2011). S2N($a$, $b$, $c$) means the node age is calibrated by a mixture of two skew-normal distributions (Wilkinson, et al. 2011). G($a$, $b$) means the node age is calibrated by a gamma distribution with shape $a$ and rate $b$.

c. Detailed justifications for these fossil calibrations are given in the appendix. For all other calibrations, justifications are in Benton, et al. (2009) and dos Reis, et al. (2012).

d. Calibration densities are the posterior distribution from a model of fossil preservation and discovery with species diversification that takes into account the effects of the K-Pg extinction in the model (Wilkinson, et al. 2011).
assuming a uniform prior on models. With the highest marginal likelihood (probability in each dataset is shown in bold type. The rate priors are \( G(2, 1) \) and \( G(2, 20) \) for mitochondrial and nuclear data correlated rates. The age of the root is fixed to one (i.e. we use a \( 'B(0.99, 1.01)' \) calibration on the substitution model is HKY+G. Marginal likelihoods are estimated by thermodynamic integration with 64 points. All introns Nuclear UTR and introns Table 5. Bayesian selection of relaxed-clock model.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>Log Marginal ( L^a )</th>
<th>( BF^{bc} )</th>
<th>( P^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial 1st</td>
<td>SC</td>
<td>-16,519.03 (0.010)</td>
<td>1.3 ( \times ) 10^{-18}</td>
<td>1.2 ( \times ) 10^{-18}</td>
</tr>
<tr>
<td>and 2nd c.p.</td>
<td>IR</td>
<td>-16,480.58 (0.021)</td>
<td>0.063</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-16,477.82 (0.035)</td>
<td>-</td>
<td>0.94</td>
</tr>
<tr>
<td>Mitochondrial 3rd c.p.</td>
<td>SC</td>
<td>-16,684.50 (0.014)</td>
<td>-</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-16,686.29 (0.043)</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-16,685.26 (0.040)</td>
<td>0.47</td>
<td>0.29</td>
</tr>
<tr>
<td>Mitochondrial RNA</td>
<td>SC</td>
<td>-7,906.85 (0.0087)</td>
<td>0.74</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-7,908.40 (0.015)</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-7,906.55 (0.023)</td>
<td>-</td>
<td>0.53</td>
</tr>
<tr>
<td>Nuclear 1st and 2nd c.p.</td>
<td>SC</td>
<td>-32,179.80 (0.0092)</td>
<td>0.0047</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-32,175.77 (0.022)</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-32,174.44 (0.032)</td>
<td>-</td>
<td>0.79</td>
</tr>
<tr>
<td>Nuclear 3rd c.p.</td>
<td>SC</td>
<td>-24,535.33 (0.012)</td>
<td>7.2 ( \times ) 10^{-12}</td>
<td>6.7 ( \times ) 10^{-12}</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-24,512.45 (0.038)</td>
<td>0.062</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-24,509.67 (0.030)</td>
<td>-</td>
<td>0.94</td>
</tr>
<tr>
<td>Nuclear UTR and introns</td>
<td>SC</td>
<td>-64,739.20 (0.016)</td>
<td>5.7 ( \times ) 10^{-4}</td>
<td>3.8 ( \times ) 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-64,732.41 (0.038)</td>
<td>0.51</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-64,731.73 (0.046)</td>
<td>-</td>
<td>0.66</td>
</tr>
<tr>
<td>All</td>
<td>SC</td>
<td>-162,684.8 (0.024)</td>
<td>2.1 ( \times ) 10^{-103}</td>
<td>2.1 ( \times ) 10^{-103}</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-162,467.0 (0.086)</td>
<td>8.4 ( \times ) 10^{-9}</td>
<td>8.4 ( \times ) 10^{-9}</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>-162,448.4 (0.15)</td>
<td>-</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Marginal likelihoods are estimated by thermodynamic integration with 64 points. The substitution model is model is HKY+G. SC: Strict clock; IR: Independent log-normal; AR: Autocorrelated rates. The age of the root is fixed to one (i.e. we use a \( 'B(0.99, 1.01)' \) calibration on the root in MCMCTree). The rate priors are \( G(2, 1) \) and \( G(2, 20) \) for mitochondrial and nuclear data respectively. The prior on \( \alpha^2 \) is \( G(1, 1) \) in all cases. The model with the highest posterior probability in each dataset is shown in bold type. a: Values in brackets are the standard errors (see Appendix 2). b: The values are the Bayes factor of the given model compared with the model with highest marginal likelihood (see Appendix 2). c: Posterior model probabilities are calculated assuming a uniform prior on models. d: The six datasets are analysed together as six partitions.
**Figure 1.** The timetree of Primates. Nodes are drawn at their posterior mean ages in millions of years ago (Ma) estimated under the autocorrelated-rates (AR) clock model and calibration strategy A. Filled dots indicate nodes calibrated with the posterior times from the 10-species tree (inset figure), and empty dots indicate nodes with fossil constraints in the 372-species tree. Horizontal bars and numbers in parenthesis represent the 95% posterior CI for the node ages. Numbers associated with branches are ML Bootstrap support values of major clades.

**Figure 2.** Strepsirrhine portion of the primate timetree (AR clock and calibration strategy A). Legend as for figure 1.

**Figure 3.** Catarrhine portion of the primate timetree (AR clock and calibration strategy A). Legend as for figure 1.

**Figure 4.** Tarsiidae and platyrrhine portion of the primate timetree (AR clock and calibration strategy A). Legend as for figure 1.

**Figure 5.** Effect of calibration strategy and relaxed-clock model. (A) Posterior time estimates under fossil calibration strategy A vs. time estimates under strategy B, for the AR clock model. (B) Posterior time estimates under the AR vs. IR clock models, for calibration strategy A.

**Figure 6.** Infinite-sites plot. Posterior CI width is plotted against mean posterior divergence times for (A) analysis under calibration strategy A and AR clock, and (B) analysis under calibration strategy B and AR clock. In both cases, black dots indicate the eight primate nodes shared between the 10-species and 372-species trees, while the grey dots represent the rest of the nodes. Solid line: regression through the origin fitted to the black dots. Dashed line: regression through the origin fitted to all the dots.
Figure 5

Figure 6
Supplementary Figure 1