JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12076

The precision of the hominid timescale estimated by relaxed clock methods

C. G. SCHRAGO & C. M. VOLOCH

Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Keywords:

hominidae;
hominoidea;
infinite sites theory;
relaxed molecular clock;
speciation.

Abstract

The chronological scenario of the evolution of hominoid primates has been thoroughly investigated since the advent of the molecular clock hypothesis. With the availability of genomic sequences for all hominid genera and other anthropoids, we may have reached the point at which the information from sequence data alone will not provide further evidence for the inference of the hominid evolution timescale. To verify this conjecture, we have compiled a genomic data set for all of the anthropoid genera. Our estimate places the *Homo*/*Pan* divergence at approximately 7.4 Ma, the *Gorilla* lineage divergence at approximately 9.7 Ma, the basal Hominidae divergence at 18.1 Ma and the basal Hominoidea divergence at 20.6 Ma. By inferring the theoretical limit distribution of posterior densities under a Bayesian framework, we show that it is unlikely that lengthier alignments or the availability of new genomic sequences will provide additional information to reduce the uncertainty associated with the divergence time estimates of the four hominid genera. A reduction of this uncertainty will be achieved only by the inclusion of more informative calibration priors.

Introduction

Since the seminal study of Sarich & Wilson (1967), the inference of divergence times of the genera of hominids has been a recurrent theme in molecular dating analyses. All of the major developments in the field of divergence time estimation have been rapidly applied to the problem of the evolutionary timescale of the great apes and other hominoids (Goodman *et al.*, 1983; Templeton, 1983; Hasegawa *et al.*, 1985; Yoder & Yang, 2000; Schrago & Russo, 2003; Hobolth *et al.*, 2007). Apart from the obvious interest in obtaining a clearer picture of the evolution of our own species, these studies were also motivated by the large amount of molecular data that are publicly available for hominids. For instance, the genome of at least one species of all of the genera of great apes has been published to date (Lander *et al.*, 2001; Venter *et al.*, 2001; Locke *et al.*, 2011; Scally *et al.*, 2012). However, one may speculate whether the molecular data available to estimate the chronology of hominid diversification is fundamentally complete. In this sense, the inference of the hominid timescale should be performed with the minimum possible stochastic error.

Recent estimates have reported a large precision associated with the ages of hominid diversification (Chatterjee *et al.*, 2009; Perelman *et al.*, 2011; Wilkinson *et al.*, 2011). In addition, Yang & Rannala (2006) found that the primate data set of approximately 150 000 bp used by Steiper *et al.* (2004) nearly approached the theoretical limit of information from sequence data. Thus, it is possible that the study of divergence times of the genera of Hominidae will be minimally altered by the inclusion of additional sequence data or even new genomes and that the uncertainty associated with the hominid timescale is now essentially determined by the uncertainty of the calibration information obtained from the fossil record. To test this hypothesis, we have compiled an ideal data set of genes that evolved statistically under rate homogeneity. We show that the data available thus far have reached the saturation of information from sequence data and that new primate genomic sequences will probably not alter the uncertainty

Correspondence: Carlos G. Schrago, Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética, CCS, A2-092, Rua Prof. Rodolpho Paulo Rocco, SN, Cidade Universitária, Rio de Janeiro, RJ, CEP: 21941-617, Brazil.
Tel.: +55 21 2562 6397, +55 21 4063 8278; fax: +55 21 4063 8278;
e-mail: carlos.schrago@gmail.com

of the timescale of the diversification of hominid genera unless they permit the inclusion of new informative calibration priors.

Materials and methods

Sequence selection and alignment

Alignments of orthologous genes of the anthropoid species with completed or partially completed genome projects were obtained from the OrthoMam database (Ranwez *et al.*, 2007). These consisted of genomic data from *Callithrix* (The Human Genome Center, Baylor College of Medicine), *Macaca* (Gibbs *et al.*, 2007), *Nomascus* (The Broad Institute of Harvard and MIT), *Pongo* (Locke *et al.*, 2011), *Gorilla* (Scally *et al.*, 2012), *Pan* (Mikkelsen *et al.*, 2005) and *Homo* (Lander *et al.*, 2001; Venter *et al.*, 2001). Fig. 1 presents the phylogenetic nomenclature used throughout this study. A total of 9335 orthologous alignments were downloaded. Of these, 249 alignments were eliminated from further analyses because they contained more than 50% indel sites. To reduce the rate variation among branches, which would amplify the variance of the divergence time estimates, we subjected the remaining alignments to a test of the molecular clock. Using PhyML 3 (Guindon & Gascuel, 2003), we inferred the phylogenetic tree for each gene independently, with the substitution model chosen by the likelihood ratio test (LRT) implemented in HyPhy (Pond *et al.*, 2005). The inferred trees were then analysed using PAML 4.5 (Yang, 2007) to estimate whether the log-likelihood of the topology enforcing the molecular clock was significantly lower than that of the topology in which the branch lengths were freely estimated (Felsenstein, 1988). Only those genes that failed to reject the molecular clock were used for the divergence time estimation, consisting of 1367 genes, with a total of 1 619 994 nucleotide sites. The two main data sets were composed of this pool of 1367 genes. The first data set, hereafter referred to as *full*, consisted of the complete alignment of the "clock-like" genes. For the second data set, only those genes that recovered the correct phylogenetic relationship of the primate species used were selected, resulting in 373 genes, with a total of 560 880 nucleotide sites. It is worth mentioning that, although the molecular clock could not be rejected for the individual genes, rate variation among genes prevents the application of the strict clock to the concatenated supermatrix. Moreover, the strict clock is a special case of the relaxed clock when rate variation among branches is null (Drummond *et al.*, 2006).

To further investigate the effect of the sequence length on the uncertainty of the divergence time estimates, we composed three groups of alignments in which subsamples with alignment lengths of approximately 1 kbp, 10 kbp and 100 kbp were created by randomly drawing genes from the "clock-like" pool. We created 10 alignments in each group; thus, we analysed 30 additional alignments composed of subsamples (10 from each of the three alignment group lengths). The precision of the estimates was measured either by calculating the standard deviation of the marginal posterior density or by the width (*w*) of the 95% highest probability density (HPD) interval of the posterior.

Lastly, to verify whether the inclusion of additional genomic sequences would provide further information on the estimates of the hominid divergence times, we constructed two reduced data sets, both excluding *Callithrix*. The first reduced set contained *Homo*, *Pan*, *Gorilla*, *Pongo*, *Nomascus* and *Macaca* sequences. In the second reduced set, we also eliminated *Nomascus* and *Gorilla*, maintaining only 4 terminals, based on the rationale that, because the *Nomascus* and *Gorilla* divergences did not contain any additional calibration information, their elimination would not affect the estimates of the ages of other hominid divergences. If valid, we would demonstrate that unless new calibration priors are used,

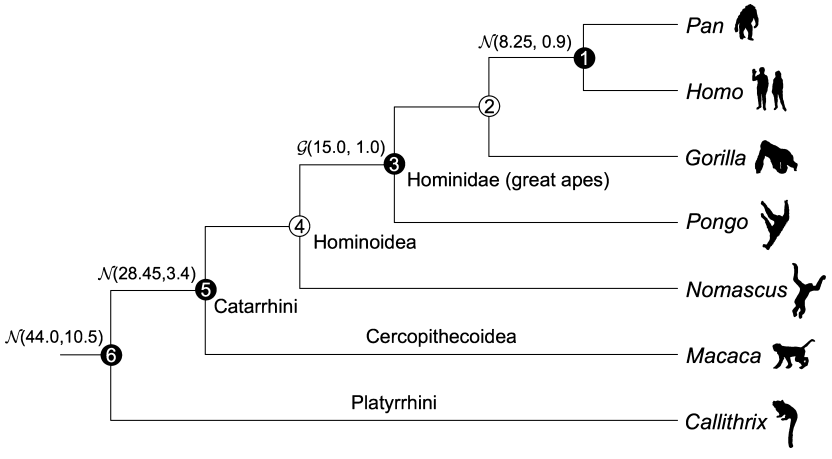

Fig. 1 Primate phylogeny illustrating the major anthropoid clades (*sensu* Groves, 2001). The nodes in which calibration information was entered are shown in black circles.

any increase in the taxonomic sampling provided by the sequencing of additional primate genomes would likely not bring any additional information regarding the timescale of hominid divergences.

Divergence time estimation and the infinite sites analysis

The divergence time estimation was conducted under a Bayesian framework using BEAST 1.7.2 (Drummond & Rambaut, 2007) and the MCMCTree program of the PAML package (Yang, 2007). All BEAST analyses were conducted using the concatenated alignment of genes under the GTR+Γ_4+I model of nucleotide substitution, which was chosen by the LRT in HyPhy. In MCMCTree, we used the HKY+G model, because it was the parametric richer model available in the program. We used the following models of evolutionary rates, as they allow for rate independence between the branches: the uncorrelated lognormal model in BEAST (Drummond *et al.*, 2006) and the independent model of MCMCTree (Rannala & Yang, 2007). Both programs implement the Markov chain Monte Carlo (MCMC) algorithm to approximate the joint posterior density of divergence times. After an adjustable period of burn-in, the Markov chains were run for 50 000 000 generations and sampled every 1000th cycle. The analyses of the MCMC output were conducted using the CODA package (Plummer *et al.*, 2006) in the R environment (www.r-project.org).

With the aim of investigating whether the posterior density of the divergence time estimates of hominoids had reached the limiting distribution when the data consisted of infinite sites, we adopted the strategy of Rannala & Yang (2007; Yang & Rannala, 2006), who derived the posterior densities of divergence times when the branch lengths parameters are considered given and fixed at their maximum likelihood (ML) estimates. Under these assumptions, the stochastic error associated with the ML estimate is eliminated, and the variance of the posterior density of node ages is exclusively determined by the uncertainty of the calibration information used. Therefore, if the precision of the posterior density is close to that predicted under the infinite sites approach, the maximum amount of information has been obtained from the molecular data. This result indicates that a further reduction of the variance of the posterior density will only be achieved by the adoption of informative calibration priors.

Calibration information

We used the calibration information described by Benton & Donoghue (2007), which places the *Homo*/*Pan* divergence from 6.5 to 10 Ma and the Cercipithecoid/Hominoid divergence from 23.0 to 33.9 Ma. These data were entered as normal priors, with the mean set at the average between the minimum and maximum values of the range and the standard deviation set to include the minimum and maximum values of the range delimiting 99% of the area under the curve. This strategy resulted in normal priors $N(8.25, 0.9)$ and $N(28.45, 3.4)$. The divergence between the orang-utan and other hominoids was calibrated by a gamma distribution with a shape parameter = 15 and a scale parameter = 1, corresponding to a distribution with a mean = 15 Ma and a 95% HPD interval from 9.2 to 21.9 Ma. This calibration was based on the fossil record of *Lufengpithecus*, *Sivapithecus* and *Khoratpithecus* from the Miocene of Asia (Hartwig, 2002; Chaimanee *et al.*, 2003, 2004). Lastly, for the root of the tree, the Platyrrhini/Catarrhini divergence was calibrated by a normal prior with a mean = 44 Ma and a standard deviation = 10.5 Ma. This distribution is centred at the average value for the divergence obtained from the timetree.org database, and the large standard deviation results in a 95% HPD interval from 26.7 Ma, which is approximately the age of the earliest platyrrhine record, *Branisella* sp. (Takai & Anaya, 1996), to 61.3 Ma. We used normal and gamma priors instead of adopting hard bounds because of the uncertainty associated with the upper and lower limits. The maximum range of a divergence time is particularly difficult to be established and several strategies were proposed to deal with this issue (Marshall, 2008; Wilkinson & Tavare, 2009; Laurin, 2012). Finally, it is worth mentioning that the node ages inferred here are based on the average genetic divergence between lineages. This is the measure commonly used by relaxed clock methods. Thus, the divergence times do not necessarily imply speciation times, that is, the cessation of gene flow between lineages (Burgess & Yang, 2008).

Results

For both of the data sets studied, the divergence times of the six primate nodes were robust with regard to the method (BEAST or MCMCTree). For the full data set, the largest discrepancy between the methods was found for the root node in which the Platyrrhini/Catarrhini divergence was dated at 51.5 Ma (39.3–63.5) in BEAST and 53.3 Ma (41.6–65.0) in MCMCTree (Table 1). However, the difference was reduced in the data set with the correct topologies only: BEAST dated the root at 50.1 Ma (36.7–62.7) and MCMCTree at 51.4 Ma (34.6–64.9) (Table 2), a statistically negligible difference. For all of the other divergences, the differences between the methods were generally less than 1 Ma.

The *Homo*/*Pan* divergence was dated at approximately 7.3 Ma using the full data set (the average between the BEAST and MCMCTree estimates). The age of the node was estimated at 7.4 Ma using the infinite sites approach. The width of the 95% HPD interval was

Table 1 Divergence times of selected primate genera using genes that failed to reject the molecular clock.

Node	Divergence	BEAST	MCMCTree	Infinite sites
1	*Homo/Pan*	7.2 (5.7–8.7)/3.0*	7.4 (5.9–8.8)/2.9*	7.4 (6.0–8.9)/2.9
2	*Gorilla*	9.4 (7.6–11.4)/3.8	9.7 (7.8–11.7)/3.9	9.7 (8.0–11.6)/3.6
3	*Pongo*	18.0 (14.6–21.4)/6.8	18.1 (14.6–21.5)/6.9	18.1 (14.7–21.7)/7.0
4	*Nomascus*	20.4 (16.5–24.2)/7.7	20.5 (16.8–24.4)/7.6	20.6 (16.8–24.6)/7.8
5	*Macaca*	27.7 (22.9–32.6)/9.7	28.2 (23.4–33.1)/9.7	28.2 (23.3–33.2)/9.9
6	*Callithrix*	51.5 (39.3–63.5)/24.2	53.3 (41.6–65.0)/23.4	53.2 (40.7–66.1)/25.4

*In Ma, the mean and 95% HPD interval.

Table 2 Divergence times of selected primate genera using genes that failed to reject the molecular clock and that presented no topological incongruence with the standard primate phylogeny.

Node	Divergence	BEAST	MCMCTree	Infinite sites
1	*Homo/Pan*	6.8 (5.4–8.4)/3.0*	7.0 (5.6–8.6)/3.0	7.0 (5.5–8.6)/3.1
2	*Gorilla*	9.8 (7.8–11.9)/4.1	10.1 (8.0–12.2)/4.2	10.1 (8.0–12.2)/4.2
3	*Pongo*	17.5 (14.2–20.7)/6.5	17.9 (14.5–21.4)/6.9	17.8 (14.3–21.3)/7.0
4	*Nomascus*	20.8 (16.9–24.6)/7.7	21.3 (17.3–25.3)/8.0	21.2 (17.2–25.2)/8.0
5	*Macaca*	29.2 (24.2–34.1)/9.9	29.7 (24.6–34.7)/10.1	29.6 (24.5–34.8)/10.3
6	*Callithrix*	50.1 (36.7–62.7)/26.0	51.4 (34.6–64.9)/30.3	52.7 (40.0–65.7)/25.7

*In Ma, the mean and 95% HPD interval.

essentially identical using the full data set, measuring 3.0 Ma in BEAST and 2.9 Ma in both MCMCTree and infinite sites (Table 1). When genes with the correct topologies were used, the age of the human/chimp divergence decreased slightly to approximately 6.9 Ma, which was very similar to the age of the node under the infinite sites model (7.0). The precision of the estimates was also measured to be approximately 3.0 Ma (Table 2). Therefore, the use of genes with correct topologies did not significantly affect either the estimation of the age of the divergence or the precision of the estimate.

The age of the node that separates the *Gorilla* and *Homo*/*Pan* lineages was inferred at 9.4 and 9.7 Ma using the full data set (Table 1), and the estimated age was estimated at 9.7 Ma using the infinite sites model. The width of the 95% HPD intervals varied from 3.9 Ma (MCMCTree) to 3.6 Ma (infinite sites). The divergence was slightly older for the correct topology data set, ranging from 9.8 Ma (the BEAST estimate) to 10.1 Ma (the estimates of both MCMCTree and infinite sites) (Table 2). The widths of the HPD intervals of the estimates were also wider, at approximately 4.2 Ma.

For the full data set, the age of the orang-utan divergence, that is, the time to the recent common ancestor (TMRCA) of the great apes, was dated at approximately 18 Ma (18.1 Ma for the infinite sites), with the width of the 95% HPD interval ranging from 6.8 Ma (BEAST) to 7.0 (infinite sites) (Table 1). This value varied from 17.5 Ma (BEAST) to 17.9 Ma (MCMCTree) when the data set with the correct topologies was used, and the infinite sites analysis measured the age of the divergence at 17.8 Ma. The precision of the estimates ranged from 6.5 Ma (BEAST) to 7.0 (infinite sites) (Table 2).

The *Nomascus* divergence, which marks the age of the TMRCA of the Hominoidea, ranged from 20.4 Ma (BEAST) to 20.6 Ma (infinite sites) in the full data set, with credibility interval widths varying from 7.6 Ma (MCMCTree) to 7.8 Ma (infinite sites) (Table 1). Using the correct topology data set, the age ranged from 20.8 Ma (BEAST) to 21.3 Ma (MCMCTree) and was 21.2 Ma using the infinite sites model. The width of the credibility intervals ranged from 7.7 Ma (BEAST) to 8.0 Ma (MCMCTree and infinite sites) (Table 2).

Outside hominoids, the Cercopithecoid/Hominoid divergence (TMRCA of extant Catarrhini primates) was dated from 27.7 Ma (BEAST) to 28.2 Ma (MCMCTree and infinite sites) for the full data set, with the width of the 95% HPD intervals varying from 9.7 Ma (BEAST and MCMCTree) to 9.9 (infinite sites) (Table 1). The age was estimated to be slightly older using the genes with the correct topology, ranging from 29.2 Ma (BEAST) to 29.7 Ma (MCMCTree), with an estimate of 29.6 Ma using the infinite sites approach. The credibility intervals were also wider, varying from 9.9 Ma (BEAST) to 10.3 Ma (infinite sites).

When the divergence times were estimated using subsamples of the original alignment, the averages of the estimates from the 1 kbp, 10 kbp and 100 kbp sampling strategies were similar to those obtained using the full data set with the MCMCTree and the infinite sites models (Fig. 2, see further information in Table S1). Regardless of the inclusion of the alignments with

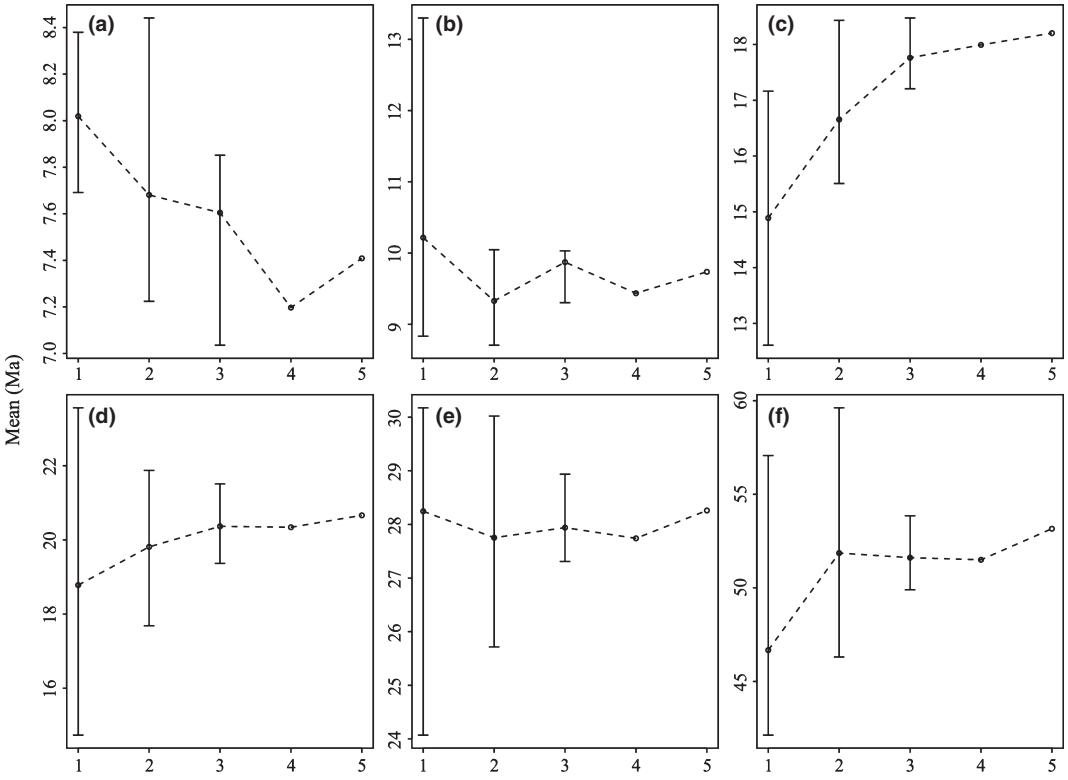

Fig. 2 Divergence times (Ma) inferred from (1) 1 kbp data sets; (2) 10 kbp data set; (3) 100 kbp data set; (4) full data set analysed using MCMCTree; and (5) infinite sites. In (1), (2) and (3), the bars depict the maximum and minimum values obtained among the 10 subsamples. (a) *Homo*/*Pan* divergence; (b) the *Gorilla* lineage divergence; (c) the *Pongo* divergence; (d) the *Nomascus* divergence; (e) the Cercopithecoid/Hominoid divergence, as represented by the *Macaca* divergence; and (f) the Catarrhini/Platyrrhini divergence.

small size, the largest discrepancies found among estimates were less than 1 Ma for the *Homo*/*Pan* (0.8 Ma), *Gorilla* (0.9 Ma) and the *Macaca* (0.5 Ma) divergences. The root node, the Catarrhini/Platyrrhini separation, resulted in the most heterogeneous estimates, with a difference of 6.5 Ma between the maximum and minimum estimates, followed by the *Pongo* (3.3 Ma) and *Nomascus* (1.9 Ma) divergences. The differences between the subsamples for all of the node ages were significantly reduced when the 100 kbp alignments were used, a pattern most dramatically depicted in the root node (Fig. 2f).

Although the means of the posterior densities were generally homogeneous, the standard deviation of the posterior densities presented a pattern of reduction from the 1 kbp data sets to the infinite sites estimates (Fig. 3). Moreover, for all of the nodes, the differences between the standard deviations (SD) were smaller between the full data set MCMCTree estimate and the infinite site estimates. This result demonstrates a tendency of stabilization (a plateau in the graph) that is most exemplified by the *Homo*/*Pan* (Fig. 3a) and the *Gorilla* (Fig. 3b) divergences for which the difference between the SDs of the MCMCTree data set and infinite sites was 0.015 and 0.011 Ma respectively. This difference was also very small for the divergences of the *Pongo* (−0.030 Ma), *Nomascus* (0.036 Ma) and *Macaca* (−0.044 Ma) lineages. The value shifted to −0.222 Ma in the root node. Likewise, the maximum and minimum SD values of the posterior density among the subsamples generally decreased from the 1 kbp to 100 kbp alignments; the only exception was the *Homo*/*Pan* divergence.

The reduction of the number of terminals demonstrated that the ages of the divergences were little affected by the presence of *Nomascus* and *Gorilla* (Table 3). The widths of the 95% HPD interval of the *Homo*/*Pan* divergence were measured as 2.9 and 3.3 Ma, with and without including *Nomascus* and *Gorilla*, respectively, and are fundamentally the same values from the full data set estimates, including the mean, which varied from 7.3 to 7.4 Ma. As in the full data set, the age of the basal hominid divergence, the *Pongo* separation, was dated at approximately 18 Ma. The *w* values ranged from 7.8 Ma (with *Nomascus* and *Gorilla*) to 8.1 (without *Nomascus* and *Gorilla*), measures that were approximately 1 Ma wider than the credibility width for the full data set.

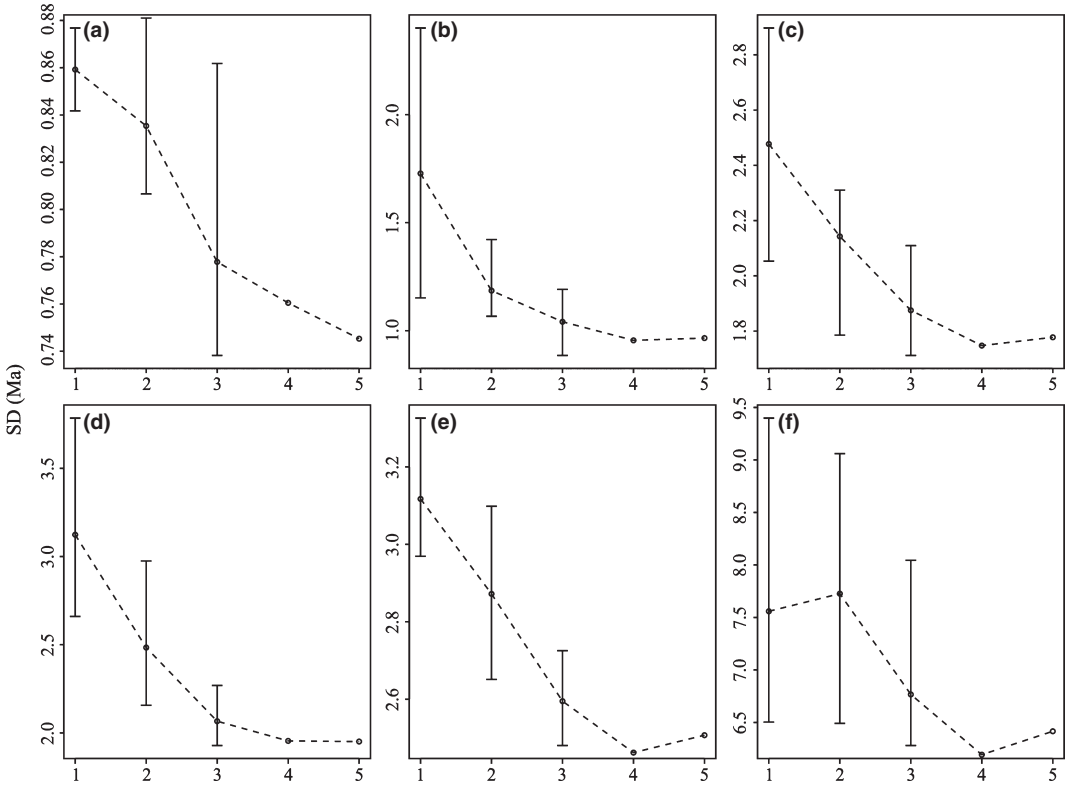

Fig. 3 Standard deviations of the posterior densities of node ages (Ma) inferred from (1) 1 kbp data sets; (2) 10 kbp data set; (3) 100 kbp data set; (4) full data set analysed with MCMCTree; and (5) infinite sites. In (1), (2) and (3), the bars depict the maximum and minimum values obtained among the 10 subsamples. (a) *Homo*/*Pan* divergence; (b) the *Gorilla* lineage divergence; (c) the *Pongo* divergence; (d) the *Nomascus* divergence; (e) the Cercopithecoid/Hominoid divergence, as represented by the *Macaca* divergence; and (f) the Catarrhini/Platyrrhini divergence.

Table 3 Divergence times of a reduced data set of primates using genes that failed to reject the molecular clock and that presented no topological incongruence with the standard primate phylogeny.

Node	Divergence	With *Nomascus* and *Gorilla*	Without *Nomascus* and *Gorilla*
1	*Homo/Pan*	7.4 (6.0–8.9)/2.9	7.3 (5.8–9.1)/3.3
2	*Gorilla*	9.8 (7.9–11.9)/4.0	NA
3	*Pongo*	18.3 (14.7–22.5)/7.8	18.5 (14.9–23.0)/8.1
4	*Nomascus*	21.0 (16.9–25.6)/8.7	NA
5	*Macaca*	29.7 (24.4–35.3)/10.9	29.6 (22.3–35.7)/13.4

*In Ma, the mean and 95% HPD interval.

Discussion

Our results indicate that the timescale of the diversification of hominid genera has reached the point of minimum uncertainty associated with the age estimates because both the 95% HPD intervals and the standard deviations of the posterior densities are very close to the value predicted by the theoretical limit (Yang & Rannala, 2006; Rannala & Yang, 2007). In addition, the inclusion of new terminals will probably have little or no effect on the precision of the divergence time estimates of the Hominidae genera. Although further precision may be achieved with more informative calibration priors, the molecular data *per se* will likely not provide any additional information. We describe the current scenario as a possibility because the methods based on the species tree reconstruction from multiallelic intraspecific data sets may shed new light on this issue (Rannala & Yang, 2003; Knowles & Carstens, 2007; Edwards, 2009; Liu *et al.*, 2009). Such data sets are likely to become usual as the cost of genomic sequencing decreases. However, these methods are

generally more parametric than the Bayesian dating analysis; thus, the variance of the estimates will possibly not decline when compared with that obtained using a large alignment of genes that failed to reject the molecular clock hypothesis.

Although the timescale of the diversification of hominid genera was inferred with minimum uncertainty, it should not be assumed that those ages refer to the speciation times of hominids. This is because our node ages were inferred from average genetic divergences, and several genes may have diverged much before the cessation of gene flow (Pamilo & Nei, 1988). To estimate speciation times, one should use another family of methods (Rannala & Yang, 2003; Degnan & Rosenberg, 2009). Nevertheless, it is likely that our conclusion still holds under those coalescent-based approaches.

It is worth noting that the uncertainty associated with the node ages of our timescale is not smaller when compared with other recent genomic estimates. For instance, Kumar & Hedges (1998) and Kumar *et al.* (2005) have proposed estimates of the *Homo*/*Pan* divergence with widths of the 95% confidence interval of only 1.96 Ma and 2.04 respectively. In fact, *w*-values as low as 1.3 Ma are found in the literature (dos Reis *et al.*, 2012). Nonetheless, a recent combined analysis of fossil modelling and molecular data rendered an estimate of *w* equal to 3.9 Ma for the *Homo*/*Pan* genetic divergence (Wilkinson *et al.*, 2011). Therefore, the uncertainty associated with hominid divergences presents some variation, which is probably caused by the method applied and the variance of the calibration priors. Here, we showed that sequence information has already reached the theoretical limit and that increased sequence lengths and taxonomic sampling will not significantly alter the hominid timescale. Accordingly, if divergence time studies applied the same methodological framework, any discrepancy among the precision of the divergence estimates would be correlated with the information from the prior distributions used to calibrate the timescale. A similar point had been raised by Kumar *et al.* (2005) and Yang & Rannala (2006), who conducted a meta-analysis of Steiper *et al.* (2004).

It is also relevant to assert that the posterior distributions of divergence times were heavily impacted by the data likelihood. For instance, in MCMCTree analysis, the 95% HPD interval of the marginal prior distribution of the age of the divergence between the *Gorilla* and *Homo*/*Pan* ranged from 7.6 to 19.0 Ma (results not shown), while the posterior ranged from 7.8 to 11.7 Ma (Table 1). Thus, the prior alone was not responsible for the width credibility intervals; the inclusion of data shifted the width of the interval from 11.4 to 3.9 Ma. Our results demonstrated that it is unlikely that this width (3.9 Ma) will be further reduced by additional sequences or even by the inclusion of new species terminals.

As indicated in the plots of the 95% HPD interval against the age of the nodes, the estimates obtained with BEAST and MCMCTree using the large alignments are close to the theoretical limit (Fig. 4a). The largest difference among the estimates is found at the root node; all of the other ages are fundamentally identical. The regression line that fitted the points of the limiting distribution resulted in the equation $w = 0.43t$, which means that, for each 10 Ma (t), approximately 4.3 Ma is added to the credibility interval (w). It is noteworthy that the MCMCTree regression line presented a smaller coefficient, 0.41, indicating that the estimates are more precise than the theoretical limit. This difference, however, is probably due to the random nature of the MCMC algorithm; in fact, two of the 100 kbp subsamples also randomly presented smaller coefficients (Fig. 4b). This same reasoning applies to the anomalous behaviour found for the data presented in Fig. 3 in which the standard deviations of the ages of some nodes are smaller than the theoretical limit (Fig. 3c, e, f).

The w x t plots also illustrated that when the sequence length decreased (from 100 to 1 kbp), the uncertainty of the estimates increased (Fig. 4b–d). For instance, the average coefficient of the 1 kbp subsamples was 0.58, and the values shifted to 0.53 and 0.48 in 10 kbp and 100 kbp subsamples respectively. As the calibration information used was identical in every case, the difference was produced by the stochastic error associated with the limited number of nucleotide sites analysed. In all cases, the fit of the regression line was significant ($p < 1\%$) and very high (average $R^2 \approx 0.98$ for all comparisons). Yang & Rannala (2006) have reported that even small alignments may result in divergence time estimates close to the theoretical limit. We showed that in the case of primate divergences, the 1 to 10 kbp alignments still resulted in estimates with large variances. However, the comparison depicted in Fig. 4a shows that the theoretical limit had been reached.

Another interesting finding is that the adoption of genes that resulted in trees with the correct topology of primates did not alter the inferred chronology of hominoid divergences. Both the mean and 95% HPD intervals were close to those estimated using the full data set (Tables 1 and 2), indicating that the uncertainty of the inferred divergence times was little influenced by the gene tree/species tree problem. Moreover, the taxon sampling did not affect the hominid divergence times. The addition of *Nomascus* and *Gorilla*, both terminals connected to calibration-free nodes, had literally no influence on the ages of the *Homo*/*Pan*, *Gorilla* and *Pongo* divergences. Nevertheless, it is worth mentioning that when compared to the full data set divergences, the TMRCA of the Hominidae (orang-utan lineage divergence) presented wider credibility intervals. Such a small discrepancy further

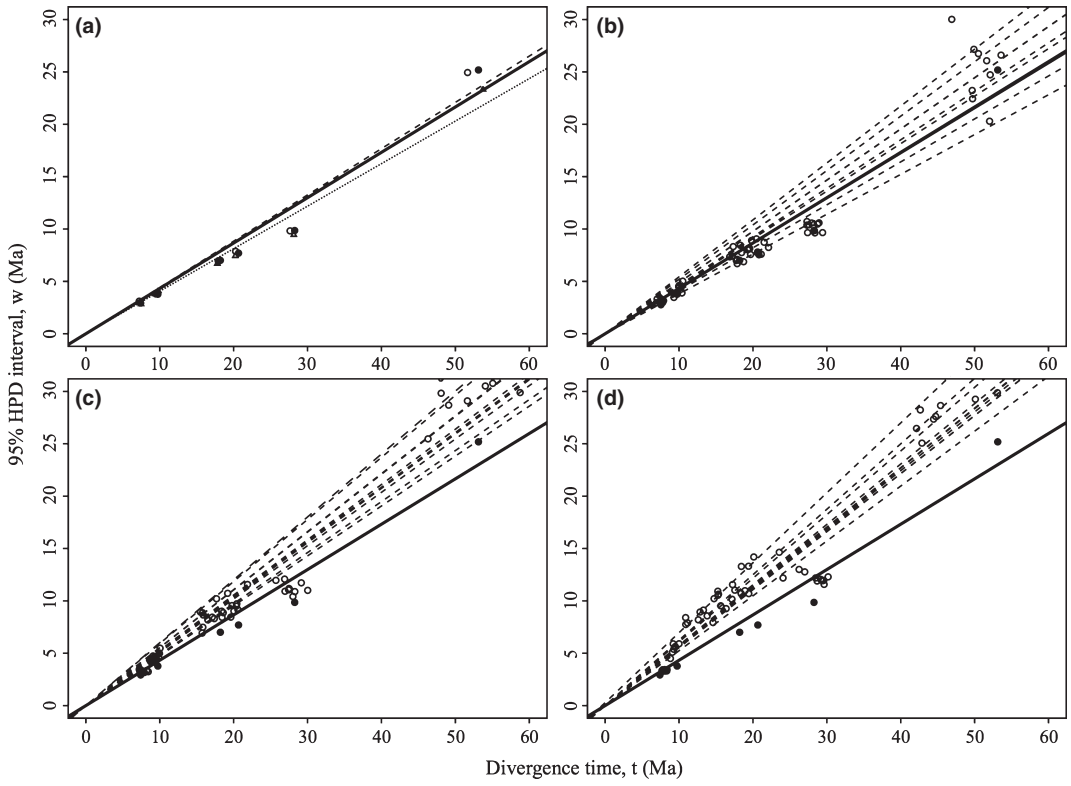

Fig. 4 Linear regressions between the widths of the 95% credibility intervals (Ma) against the node age (Ma). (a) Full data set. The solid line and solid black circles represent the theoretical limits. The dashed lines and open circles depict the estimates from the BEAST analysis. The dotted line and triangles depict the estimates using MCMCTree. (b - d) 100 kbp to 1 kbp data sets. The solid lines and solid black circles represent theoretical limits. The dashed lines and open circles depict the estimates from the 10 subsamples, with alignments of approximately 100 (a), 10 (b) and 1 (d) kbp.

confirms our conclusions because the difference was caused by the exclusion of *Callithrix*, a terminal that, unlike *Nomascus*, added additional calibration information.

Finally, the divergences presented here are in agreement with the fossil record of anthropoid primates; our estimates are also equivalent when compared with the works published recently. For example, the exhaustive analysis, with large species sampling conducted by Perelman *et al.* (2011), dated the *Homo*/*Pan* divergence at 6.6 Ma, whereas we inferred it at approximately 7.3 Ma. When the credibility interval is taken into account, the divergences are statistically equivalent. Even the study of dos Reis *et al.* (2012), which presented very precise estimates for mammalian divergences, is not at odds with the our hominoid chronological scale. The Platyrrhini/Catarrhini divergence, the root node, is an exception. Our estimates placed this divergence around 50 Ma, approximately 10 Ma older than recent estimates, varying from 37.7 (dos Reis *et al.*, 2012) to 43.5 Ma (Perelman *et al.*, 2011). A great part of such discrepancies may be due to the modelling applied to evolutionary rate at the root node, together with the large standard deviation of the prior distribution.

In conclusion, our study has demonstrated that the available sequence data from genome projects have achieved the saturation of information regarding the estimates of the divergence times of the four genera of the Hominidae. The differences among studies in the measures of uncertainty associated with the inferred ages are likely due to the set of calibration priors used because even taxon sampling played a minor role when no calibration prior was added to the data set. Therefore, to obtain a more precise chronology of hominid divergences, developments in methodological approaches or modelling strategies are required. Although a significant effort has been directed to genome projects over the last decade, chronological inference from molecular data heavily depends on paleontological research and on the availability of a good fossil record. After reaching saturation of information from sequence data, the adoption of informative calibration priors is the most effective strategy to reduce the credibility intervals of divergence times.

Acknowledgments

This work was funded by the Brazilian Research Council (CNPq) grant 308147/2009-0 and FAPERJ grants E-26/103.136/2008, 110.838/2010, 110.028/2011 and 111.831/2011 to CGS.

References

Benton, M.J. & Donoghue, P.C. 2007. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**: 26–53.

Burgess, R. & Yang, Z. 2008. Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* **25**: 1979–1994.

Chaimanee, Y., Jolly, D., Benammi, M., Tafforeau, P., Duzer, D., Moussa, I. *et al.* 2003. A Middle Miocene hominoid from Thailand and orangutan origins. *Nature* **422**: 61–65.

Chaimanee, Y., Suteethorn, V., Jintasakul, P., Vidthayanon, C., Marandat, B. & Jaeger, J.J. 2004. A new orang-utan relative from the Late Miocene of Thailand. *Nature* **427**: 439–441.

Chatterjee, H.J., Ho, S.Y., Barnes, I. & Groves, C. 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol. Biol.* **9**: 259.

Degnan, J.H. & Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**: 332–340.

Drummond, A.J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**: 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**: 699–710.

Edwards, S.V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* **63**: 1–19.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**: 521–565.

Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R. *et al.* 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222–234.

Goodman, M., Braunitzer, G., Stangl, A. & Schrank, B. 1983. Evidence on Human Origins from Hemoglobins of African Apes. *Nature* **303**: 546–548.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**: 696–704.

Hartwig, W.C. 2002 *The Primate Fossil Record*, 1st edn. Cambridge University Press.

Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.

Hobolth, A., Christensen, O.F., Mailund, T. & Schierup, M.H. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet.* **3**: 294–304.

Knowles, L.L. & Carstens, B.C. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* **56**: 887–895.

Kumar, S. & Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917–920.

Kumar, S., Filipski, A., Swarna, V., Walker, A. & Hedges, S.B. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl Acad. Sci. USA* **102**: 18842–18847.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J. *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Laurin, M. 2012. Recent progress in paleontological methods for dating the Tree of Life. *Front. Genet.* **3**: 130.

Liu, L., Yu, L.L., Pearl, D.K. & Edwards, S.V. 2009. Estimating Species Phylogenies Using Coalescence Times among Sequences. *Syst. Biol.* **58**: 468–477.

Locke, D.P., Hillier, L.W., Warren, W.C., Worley, K.C., Nazareth, L.V., Muzny, D.M. *et al.* 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* **469**: 529–533.

Marshall, C.R. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *Am. Nat.* **171**: 726–742.

Mikkelsen, T.S., Hillier, L.W., Eichler, E.E., Zody, M.C., Jaffe, D.B., Yang, S.-P. *et al.* 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.

Pamilo, P. & Nei, M. 1988. Relationship between gene trees and species trees. *Mol. Biol. Evol.* **5**: 568–583.

Perelman, P., Johnson, W.E., Roos, C., Seuanez, H.N., Horvath, J.E., Moreira, M.A. *et al.* 2011. A molecular phylogeny of living primates. *PLoS Genet.* **7**: e1001342.

Plummer, M., Nicky, B., Cowles, K. & Vines, K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **6**: 7–11.

Pond, S.L.K., Frost, S.D.W. & Muse, S.V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.

Rannala, B. & Yang, Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**: 1645–1656.

Rannala, B. & Yang, Z.H. 2007. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**: 453–466.

Ranwez, V., Delsuc, F., Ranwez, S., Belkhir, K., Tilak, M.K. & Douzery, E.J. 2007. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* **7**: 241.

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R.J., Donoghue, P.C. & Yang, Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**: 3491–3500.

Sarich, V.M. & Wilson, A.C. 1967. Immunological Time Scale for Hominid Evolution. *Science* **158**: 1200–1203.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J. *et al.* 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**: 169–175.

Schrago, C.G. & Russo, C.A. 2003. Timing the origin of New World monkeys. *Mol. Biol. Evol.* **20**: 1620–1625.

Steiper, M.E., Young, N.M. & Sukarna, T.Y. 2004. Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl Acad. Sci. USA* **101**: 17021–17026.

Takai, M. & Anaya, F. 1996. New specimens of the oldest fossil platyrrhine, Branisella boliviana, from Salla, Bolivia. *Am. J. Phys. Anthropol.* **99**: 301–317.

Templeton, A.R. 1983. Phylogenetic Inference from Restriction Endonuclease Cleavage Site Maps with Particular Reference

to the Evolution of Humans and the Apes. *Evolution* **37**: 221–244.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G. *et al.* 2001. The sequence of the human genome. *Science* **291**: 1304–1351.

Wilkinson, R.D. & Tavare, S. 2009. Estimating primate divergence times by using conditioned birth-and-death processes. *Theor. Popul. Biol.* **75**: 278–285.

Wilkinson, R.D., Steiper, M.E., Soligo, C., Martin, R.D., Yang, Z.H. & Tavare, S. 2011. Dating Primate Divergences through an Integrated Analysis of Palaeontological and Molecular Data. *Syst. Biol.* **60**: 16–31.

Yang, Z.H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.

Yang, Z.H. & Rannala, B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**: 212–226.

Yoder, A.D. & Yang, Z. 2000. Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* **17**: 1081–1090.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Divergence times of primates using genes that failed to reject the molecular clock.

Received 14 September 2012; revised 14 November 2012; accepted 15 November 2012