

Syst. Biol. 58(3):348–360, 2009
Copyright © Society of Systematic Biologists
DOI:10.1093/sysbio/syp028
Advance Access publication on July 1, 2009

Time Dependency of Molecular Rates in Ancient DNA Data Sets, A Sampling Artifact?

REGIS DEBRUYNE* AND HENDRIK N. POINAR

Ancient DNA Centre, Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L9, Canada;

**Correspondence to be sent to: Ancient DNA Centre, Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L9, Canada; E-mail: debruyne@mcmaster.ca.*

It is common knowledge that the instantaneous rate of mutation (RoM) in DNA sequences exceeds the long-term rate of substitution (RoS) when measured in interspecific phylogenetic analyses. The neutral theory of molecular evolution describes this temporary excess diversity as transient polymorphisms either removed from the population through the actions of purifying selection or fixed by random genetic drift over a few generations (Kimura 1983). Observations of these “accelerations” in the molecular rates within recent evolutionary time have been documented (Parsons et al. 1997; Lambert et al. 2002); however, they did not resolve the magnitude and duration of this phenomenon. Howell et al. (2003) have addressed these issues through pedigree analyses of human mitochondrial (mt) hypervariable region (HVR) sequences and have suggested a 5- to 10-fold acceleration compared with the long-term RoS. In addition, BurrIDGE et al. (2008) have shown that the calibration of the mt clock for galaxiid fishes using geological divergence dates with cytochrome *b* and control region sequences supports a transition period during which the RoM would decrease toward the RoS extending up to ~200 kyr (BurrIDGE et al. 2008). However, the general applicability of these specific results remains untested.

The recent publication (Ho, Phillips, Cooper, et al. 2005) of a Bayesian analysis supporting that a drastic acceleration in the molecular rates as time approaches 0 is a general and predictable feature has become a hotly debated topic in both systematics and evolutionary biology (Bandelt 2007; Emerson 2007; Ho, Kolokotronis, et al. 2007; Ho, Shapiro, et al. 2007; Howell et al. 2008). When using the program BEAST (Drummond and Rambaut 2007), with data sets containing sequences of

vertebrate taxa in conjunction with a wide range of calibration points, Ho, Phillips, Cooper, et al. (2005) have proposed a “time-dependency” model in which the rate reflects a direct response of the divergence time between the terminals. In this model, the molecular rate, referred to as the rate of change (RoC), is allowed to decrease rapidly along a vertically translated exponential decay curve until the long-term RoS reaches equilibrium. Ho, Phillips, Cooper, et al. (2005) have evaluated the putative effects of sequencing and calibration errors as well as mutational saturation in their model and have concluded that purifying selection was the most likely contributing factor involved in the variation of the RoC over time.

This model, however, describes a phenomenon an order of magnitude beyond all previous reports of such accelerations, as it allows the RoC to vary over a wide range of rates (up to and beyond 20-fold) during an extended period of time (up to 2 million years). Although intuitively appealing and explaining the data analyzed by Ho, Phillips, Cooper, et al. (2005) well, this model supports a serious and prolonged impact of deleterious mutations and would thus require a few adjustments to the current evolutionary paradigm of genomes (Penny 2005).

Recently, Bandelt (2007) and Emerson (2007) have questioned both the model of the time dependency and the significance of the rate acceleration phenomenon. Emerson (2007) has emphasized the critical role of the selection of priors in BEAST analyses and has shown that the results described by Ho, Phillips, Cooper, et al. (2005) could only be retrieved under specific conditions within a full Bayesian framework (where the level of enforcement over the priors is minimal). In this paper, our primary objective was thus to re-address the nature

of the causal factor(s) of the rate acceleration described by Ho, Phillips, Cooper, et al. (2005) as well as their biological meaning. Based on previously published material, we suggest that the emphasis placed on the divergence time in the current explanation of this phenomenon might have hidden other relevant factors such as the information content of the data sets. In order to compare the performance of the strict “time-dependency” model with a more inclusive “signal-dependency” hypothesis, we examine the impact of sequence length over the estimates of the RoC for 2 different calibrations. We show that our hypothesis for a signal-dependent artifact appears to model the data presented here more accurately and may explain some inconsistencies between published reports on evolutionary rates.

EVIDENCE SUPPORTING THE PHENOMENON OF RATE ACCELERATION

In their original paper, Ho, Phillips, Cooper, et al. (2005) reported an apparent acceleration of the RoC in recent times (<2 Ma) for 3 groups of mt data sets calibrated with dates ranging from 29 to 42 ka for tip ^{14}C calibrations to 0.125–35 Ma for node (internal) calibrations using the software application BEAST. This software, developed by Drummond and Rambaut (2007), simultaneously analyzes genetic data with their associated dates/ages of terminals and/or nodes, in order to infer, under various clock models within a Bayesian framework, the substitution rates, phylogenetic structure, branch lengths, and demographic parameters, which best describe the data at hand. In his re-analysis of the same data sets, Emerson (2007) has documented evidence of 3 main sources of error: 1) the accuracy of the phylogenetic methodology to estimate distances/branch lengths, 2) the quality of the molecular sampling, and 3) the accuracy of the time calibrations. He has shown that when different sets of priors were selected in BEAST, the exponential decay rate as seen by Ho, Phillips, Cooper, et al. (2005) could either not be retrieved (for the data sets using primates control region and protein-coding sequences) or not with the same magnitude (for data sets of protein-coding genes in avian taxa). Of the original data set, only the human Neandertal HVR sequences (8 sequences of 356 bp with 34 informative characters) produced a consistently elevated RoC in both papers when the calibration method used was based exclusively on the ^{14}C radiocarbon dates of the 4 Neandertal sequences. However, Emerson (2007) observed no rate acceleration when the time of the most recent common ancestor (tMRCA) of the Neandertal sequences was implemented as a node calibration. These results exemplify how the rate acceleration itself appears to be heavily dependent on the analytical framework of BEAST analyses. Ho, Phillips, Cooper, et al. (2005) advocate for a “full Bayesian framework” with parameter-rich models and unconstrained priors, whereas Emerson (2007) and Bandelt (2007) rather pro-

mote more tightly controlled priors and calibrations within simple models.

Within a full Bayesian framework, Ho, Kolokotronis, et al. (2007) have consistently documented elevated RoCs for BEAST analyses based on 19 data sets including ancient DNA (aDNA) sequences and calibrations (Fig. 1) except for a *Chlorobium* data set (which also had the oldest calibration point at 206 ka). These analyses shared striking similarities with the original human Neandertal comparison: 1) all were derived from relatively short, potentially low informative, sequences (range 114–741 bp; 350 bp average), 2) all were performed using time calibration priors that relied exclusively on ^{14}C radiocarbon dates of the terminals, and 3) all yielded elevated RoCs associated with widely distributed highest posterior densities (HPD). We will explore below the reasons why these conditions might be inseparable from the rate acceleration phenomenon described by Ho, Phillips, Cooper, et al. (2005).

Inconsistencies in the Rate Estimates between Simulated and Real Data Sets

Although much recent literature confirms the repeated occurrence of the rate acceleration, Emerson (2007) has argued that this result might be artificial. In response to this criticism, Ho, Kolokotronis, et al. (2007) have performed BEAST analyses on simulated data sets. The rationale for this analysis was that if BEAST yielded accurate posterior estimates of the RoC that had been used as a prior to generate the simulated data sets in a full Bayesian framework (using recent tip calibration

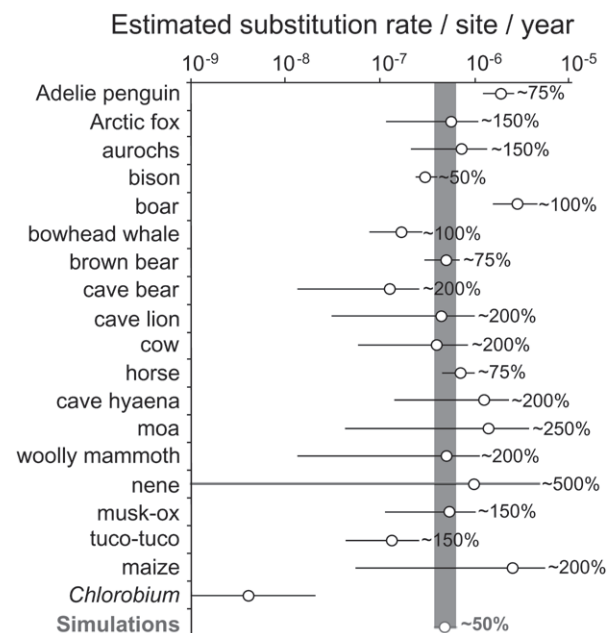


FIGURE 1. Comparison of the RoC estimate through BEAST analyses using real (in black) or simulated data sets (in gray). The width of the 95% HPD for the simulated data sets is shaded in order to compare with the real data sets. Figure modified after the figure 1 of Ho, Kolokotronis, et al. (2007, p. 3).

only), it would symmetrically confirm the accuracy of the posterior rate estimates recovered for the real data. However, for those comparisons to be relevant, the data and results from the simulated and real data sets should be as similar as possible.

Ho, Kolokotronis, et al. (2007) generated nucleotide sequences of 1 kb in length representing 30 000 years of variation with a RoS as high as 5×10^{-7} per site per year (about 25-fold the estimate of the substitution rate for the mt genome of vertebrates). With such a high RoS, one would expect on average 1 change per sequence every 2000 years, so that a representative picture of their divergence could be captured within radiocarbon time. For these simulated data sets, the estimate of the posterior RoC using BEAST was not only accurate but also extremely precise. The precision defined the width of the HPD for the RoC, which generally spanned about 50% of the average estimate (95% HPD typically spanning $\sim 2.5 \times 10^{-7}$ substitution per site per year for a mean RoC of $\sim 5 \times 10^{-7}$).

However, the precision of the posterior RoC for the real aDNA mt data sets analyzed in the same paper was generally much lower (Ho, Kolokotronis, et al. 2007): with the exception of the bison alignment, the HPDs for the rates were on average >3 times wider than for the simulations (range: 75–500% of the mean RoC; Fig. 1). Thus, despite their close average posterior RoCs, the simulated data sets do not mimic the behavior of most real aDNA data sets in which the elevated posterior rate is associated with a wide uncertainty (which could be the result of limited informativeness). One outstanding question remains to be addressed: what would the accuracy and precision of the posterior RoC be if a slower RoS, in the range of interspecific mt substitution rates (between 1 and 2×10^{-8} substitution per site and per year), were applied to simulate the same sequence data? Would one recover the expected rate through the Bayesian inference or rather an accelerated RoC with a wide HPD (or anything else)? Until those verifications are performed, it seems premature to consider that the simulated data sets are equivalent to real aDNA-calibrated data sets. Thus, if the actual acceleration in the estimates of the average RoC in the real data sets analyzed by Ho, Kolokotronis, et al. (2007) seems indisputable, it is unsure whether the simulations have demonstrated that it is not the result of an artifact.

The Causes of Rate Acceleration are Ambiguous

To help illuminate possible causes of the rate acceleration, we must carefully examine the analysis of Ho, Shapiro, et al. (2007) of an ancient bison data set (Shapiro et al. 2004) in which the authors pooled sequences according to increasing age (from 0–10 to 0–60 ka). In agreement with their model, they correlated the decrease in time depth with an increase in the posterior RoC. This correlation, however, is only visible if one focuses on the average estimate of the RoC: the analysis of its HPD rather shows that the precision of the estimate is drastically reduced when the calibration depth de-

creases (Fig. 2). Rather than a correlation between time depth and molecular rate, one might instead suggest a correlation between time depth and level of uncertainty on the posterior RoC.

We were also able to fit an exponential correlation between the calibration depth and the level of differentiation between the sequences (as estimated by the number of variable positions), supporting the expected decrease in the information content of the data set when the divergence time tends toward 0 (Fig. 2). Taken together these observations support a positive relationship between the amount of sequence variation and the precision (if not the accuracy) of the posterior estimates of the RoC. This was acknowledged by Ho, Shapiro, et al. (2007, p. 518): “Some component of the pattern could be due to a bias because as the calibrations move closer to the present, there is a decreasing amount of information in the sequences. Consequently, there is greater uncertainty on the inferred rates, so that the posterior distribution spreads out and the mean estimate increases.” However, they claim that “the magnitude of the estimation bias is not sufficient to explain the rate patterns obtained in our study” (Ho, Shapiro, et al. 2007, p. 519). Figure 2 calls this into question as the HPDs overlap extensively throughout the time range analyzed.

Whereas Ho, Phillips, Cooper, et al. (2005; Ho, Kolokotronis, et al. 2007; Ho, Shapiro, et al. 2007) have assumed that there was an unambiguous link between the divergence time and the acceleration of the average RoC, the published material certainly calls for more circumspection to characterize that phenomenon. The variation in the range of the HPDs suggests that several inter-related factors might be responsible for that phenomenon rather than a strict time dependency of the rate.

THE SIGNAL-DEPENDENCY HYPOTHESIS

It has been previously pointed out that the information content of the published data sets related to the time-dependency hypothesis might be limited. Ho, Phillips, Cooper, et al. (2005) themselves have acknowledged “large credibility intervals on the rate estimates obtained from recent calibration points,” which they tied to “low sequence variation in these alignments” (Ho, Phillips, Cooper, et al. 2005, p. 1563). Bandelt (2007) has considered the original human HVR data far too uninformative to resolve the time scales analyzed. Emerson (2007) noted possible topological issues affecting BEAST due to a lack of phylogenetic signal in the human Neandertal HVR data set. However, the impact of the information content of the data sets in the rate acceleration phenomenon has never been formally studied to date. Here, we propose an analytical framework to evaluate the relevance of a “signal-dependency” hypothesis, which assumes that the accuracy and precision of the RoC estimates (and related dates) are generally affected by the extent of sequence divergence rather than the calibration depth only.

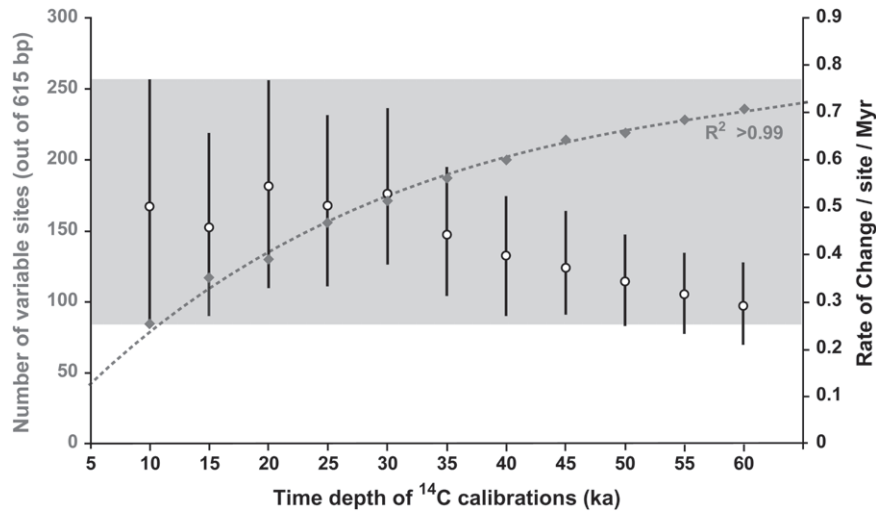


FIGURE 2. Comparisons of the time depth of the bison data set with the amount of variable sites and the posterior RoC using data derived from table 2 in Ho, Shapiro, et al. (2007). Time depth is shown in 1000 years before present. An exponential association of the amount of variable sites with time (filled diamonds, left scale) was fitted ($y = 263.77(1 - e \times 10^{-0.04x})$; R^2 displayed). Average (empty circles) and 95% HPD (bars) are displayed for the RoC (right scale). The HPD derived from the analysis of the last 10 000 years was shaded to show the overlap with the RoCs derived from the more inclusive time frames.

Factors Affecting the Information Content in Molecular Data Sets

The phylogenetic signal in a molecular data set is directly dependent on the divergence between sequences, to which we will refer to here as the amount of change (AoC), by analogy with the terminology of the rates. For a fixed number of terminal sequences, the AoC is linearly and positively correlated with 3 factors: the RoC, the sequence length, and the evolutionary time. Thus, for a known RoC, the average AoC should be the same in a 1-kb DNA sequence after 10 000 years as in a 10-kb sequence after 1000 years. This correlation deviates from linearity when mutational saturation occurs, a situation unlikely in the aDNA data sets previously analyzed (Ho, Kolokotronis, et al. 2007), given their limited amount of variable positions (except if a handful of mutational hot spots would concentrate all the sequence divergence).

For a set of known sequences, it is thus possible to alter directly the AoC in 2 symmetrical approaches: 1) one can, as in the demonstrations of the rate acceleration by Ho, Phillips, Cooper, et al. (2005; Ho, Shapiro, et al. 2007), vary the calibration depth or 2) rather vary the sequence length. Strikingly, whereas the effect of calibration time has been largely analyzed, the effect of sequence length over the posterior RoC has been completely neglected.

Predictions of the Alternate Hypotheses

To compare the performance of the time-dependency and the new signal-dependency hypotheses with real data, our analyses have been placed in a 3-dimensional framework where the molecular rate is allowed to vary along the divergence time and fragment length (Fig. 3). In the time-dependency model, any variation of frag-

ment length should have no measurable effect on the RoC (Fig. 3a). However, if the rate is generally affected by the AoC between sequences, it should follow a parallel trend when both factors vary (Fig. 3b): the posterior RoC should decrease rapidly when sequence length increases before stabilizing around a rate comparable to the interspecific RoS, as it symmetrically does when the calibration depth of the data set increases (Ho, Phillips, Cooper, et al. 2005). By analyzing the variation of the posterior RoC according to sequence length, one should be able to select which hypothesis of either the time dependency or the signal dependency best describes the data and if either of those can explain all the components of the variation of the RoC (in average and HPD).

TESTING THE HYPOTHESES

Data Analyzed

Until recently, it has been difficult to investigate the effect of sequence length on actual aDNA sequences due to their limited length (i.e., typically shorter than 1 kb and commonly below 500 bp; see Ho, Kolokotronis, et al. [2007] for a review). The recent publication of 11 radiocarbon-dated woolly mammoth mt genomes (Krause et al. 2006; Poinar et al. 2006; Rogaev et al. 2006; Gilbert et al. 2007) has provided a much longer data set (over 16 kb) of aDNA sequences. It is placed into phylogenetic context (Fig. 4) with the mt genomes of 2 modern African elephants (Hauf et al. 2000; Rogaev et al. 2006), 1 Asian elephant (Rogaev et al. 2006), and 1 Late Pleistocene American mastodon (age estimate 90 ka; Rohland et al. 2007).

Sequence alignments were performed using the CLUSTALX algorithm (as implemented in BIOEDIT ver.7.0.5; Hall 1999) with default parameters and yielding no ambiguous positions. Hexanucleotide repeats

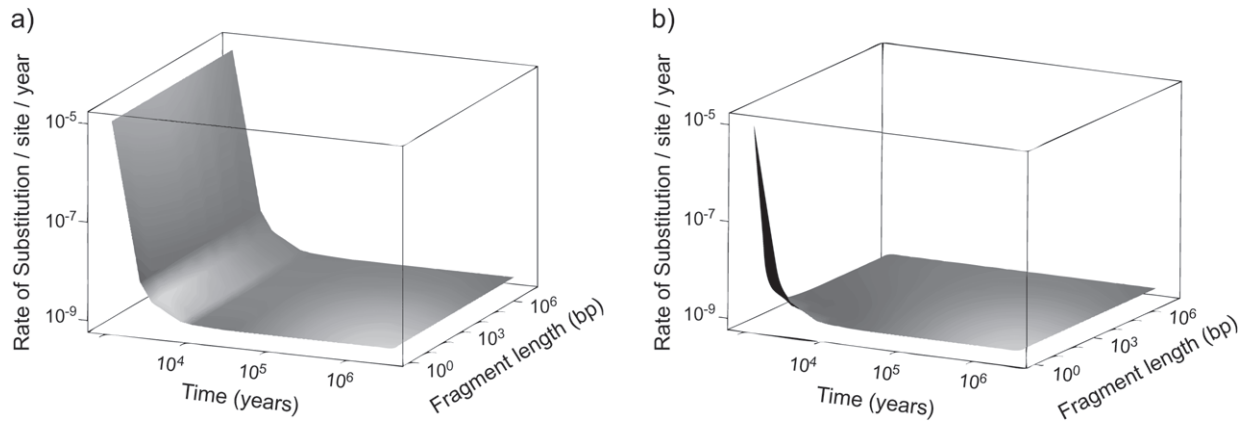


FIGURE 3. Theoretical predictions of the 2 alternate hypotheses to explain the variation for the posterior RoC reported with BEAST. (a) Ho, Phillips, Cooper, et al. (2005)'s model where the substitution rate is a variable only related to the time depth of calibration. (b) The present hypothesis of a signal-dependent artifact where the substitution rate is related to the total AOCs in sequences.

(250–450 bp long) found between the central conserved region and HVR II of the control region of Elephantidae were removed. The remaining portion of the genome (16 469 bp total) was analyzed as a single partition. In order to analyze the impact of sequence length, subsets of the mt genome alignment were generated spanning 1%, 2%, 5%, 10%, and up to 100% (per 10% increment) of the 16 469 bp. Although the haploid mt genome is expected to evolve as a single molecular entity with extremely low recombination levels in mammals, its genetic characterization and phylogenetic content are variable along its entire sequence. In order to minimize the impact of incomplete sampling of the genome, we generated 10 aligned subsets of each length by random resampling using the command SEQBOOT within the PHYLIP package (version 3.6; Felsenstein 2005). Analyses of the variation of the RoC through the sequential addition of pseudoreplicates (from 1 to 10) suggested that artifacts due to unrepresentative sampling of the genome were unlikely if >5 pseudoreplicates were analyzed for each fragment length (data not shown).

Selecting Priors for Analysis

Each of the 121 data sets was evaluated individually using the Akaike information criterion of MODELTEST (ver.3.7; Posada and Crandall 1998) to identify the most suitable substitution model. Variation in the model selected was observed with model complexity typically ranging from HKY + I + G to GTR + I + G. To limit the impact of individual models over the variation of the RoC, all Bayesian analyses were performed under the HKY + I + G model. Independent analyses of a single pseudoreplicate of each fragment length under the GTR + I + G yielded extremely close RoC estimates and suggest the independency of the pattern of rate variation over fragment length from the substitution model implemented (data not shown). In their simulations, Ho, Kolokotronis, et al. (2007) have shown that the demographic model selected in BEAST had little impact on

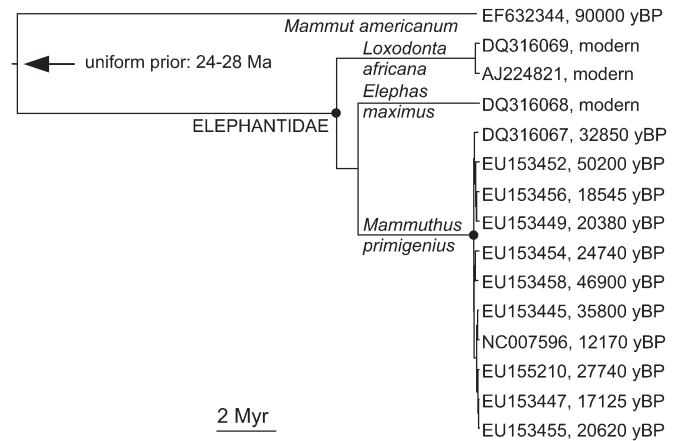


FIGURE 4. Full mt maximum credibility tree (from the shallow calibration) of the proboscidean data set. Termini accession numbers in Genbank and ^{14}C dates, used as calibration points in both sets of analyses, are indicated on tips. The uniform prior used in the second set of analysis is indicated on the root. The 2 nodes investigated for tMRCA are highlighted with bullets.

the inferred RoC. In order to minimize the number of nuisance parameters when determining the RoC, our analyses were conducted using a constant-size demographic model.

Bayes factor comparisons of the analyses of the full mt data set (using TRACER ver1.4; Rambaut and Drummond 2007) revealed a negligible advantage (in terms of benefit in marginal likelihood) of the uncorrelated lognormal relaxed clock model over the strict clock ($\log_{10} \text{BF} < 1$). Furthermore, the analysis performed under the uncorrelated lognormal relaxed clock prior yielded a low level of variation of the rate along branches (uncorrelated standard deviation median/mean = 0.50/0.58). Analyses of a single pseudoreplicate per length under a lognormal uncorrelated relaxed clock prior recovered the same pattern of rate variation as the strict clock with the predicted exception of consistently wider HPDs (data not shown). Although

potentially not as accurate as a relaxed clock model (Ho, Phillips, Drummond, et al. 2005), the strict clock model was thus preferred in our large-scale analysis of the fragment length.

Shallow (Tip) versus Deep (Tip + Node) Calibration Methods

Two sets of Bayesian analyses, which differed only by the calibration depth of the strict clock, were conducted in parallel for all data subsets (i.e., 121 individual analyses in each calibration set). These calibration methods (referred to here as shallow and deep) were performed to analyze the conditions where any rate acceleration might be observed. In the first set, we limited our calibration to the age of the 15 terminals (^{14}C dates or modern; Fig. 4). In addition to those tip dates, we implemented a uniform prior on the age of the root at 24–28 Ma for the second set of analyses. This time range corresponds to the current estimate of the divergence between the lineage of the mammutids (which includes the American mastodon) and the elephantids (Fig. 4) and is fairly well constrained by paleontological evidence (Shoshani et al. 2006).

Bayesian Analyses Using BEAST

Within each calibration set, each pseudoreplicate (for every fragment length) was analyzed separately with BEAST (ver1.4.6; Drummond and Rambaut 2007) in order to derive a single posterior RoC (mean and associated HPD) for the most credible tree. Thereafter, the 10 individual pseudoreplicates of each length were combined using TRACER to derive 1 average RoC (with associated HPD) per fragment length. All values were compared with the statistics obtained when using the entire 16 469 bp (referred to as “full mt”). This analytical protocol is similar to the one developed by Ho, Phillips, Cooper, et al. (2005; Ho, Shapiro, et al. 2007) and implemented (with tighter priors) by Emerson (2007) to address the rate acceleration where the variable parameter in our analyses was the fragment length rather than the calibration depth.

Analyses were performed for 10 000 000 generations within the full Bayesian framework advocated by Ho, Shapiro, et al. (2007). Sampling for all parameters occurred every 1000 generations after a burnin period of 1 000 000. TRACER was used to check for stationarity as well as sufficient sampling of priors ($\text{ESS} > 100$) as recommended (Drummond and Rambaut 2007). The posterior distribution of 3 parameters was analyzed: the RoC (i.e., clock rate in BEAST), as well as the age of 2 undisputed nodes that refer to the tMRCA of the Elephantinae, and the tMRCA of our *Mammuthus* sequences (Fig. 4). The monophyly of those nodes was not enforced but their posterior probability (PP) was equal to 1.0 in both the full mt and the shorter analyses down to 5% of the original length. In the shortest analyses (below 5%), the PP drastically decreased when using the shallow calibration, with the node *Mammuthus* not always recovered as monophyletic (a topological limita-

tion of the full Bayesian framework previously outlined by Emerson 2007).

PATTERN OF VARIATION OF THE ROC

Rate Variation under Shallow or Deep Calibration

Figure 5 summarizes the results of the Bayesian analyses for the posterior distribution of the RoC over the range of sequence lengths (from 1% to full mt) for both shallow (in blue) and deep (in orange) calibration methods. This figure shows that, under conditions comparable with the ones which have produced accelerated rates for other aDNA material in the past (i.e., for short sequences with shallow radiocarbon calibration), the proboscidean data also yield an elevated RoC characterized by a very wide HPD. Indeed, for fragments spanning less than 1 kb (1–5% of the total genome; Fig. 5), the shallow calibration method returns average posterior RoCs that vary from 10- to 100-fold the RoS published by Rohland et al. (2007) for the Elephantoidae (4.2×10^{-9}). These posterior RoCs are also extremely imprecise, with HPDs spanning between 250 and 400% of their average estimate and therefore comparable with the RoC estimate proposed by Ho, Kolokotronis, et al. (2007) for a data set of woolly mammoth mt HVR of similar length (741 bp; Fig. 1). Conversely, when the deep calibration is used, the sequence lengths <1 kb consistently return relatively imprecise RoCs (HPDs span: 130–170% of the mean) but which consistently overlap with the Elephantoidae’s RoS (Fig. 5).

These results confirm the expected pattern of rate acceleration when the calibration depth approaches 0, with the same modalities as previously observed. However, when compared with the predictions based on the 2 alternatives shown in Figure 3, the same observation does not specifically support the time-dependency hypothesis nor the signal-dependency hypothesis. Therefore, only the pattern of variation of the RoC with fragment length can help distinguishing between the 2.

Rate Variation over Fragment Length

The increase in fragment length has almost no effect on the average RoC derived from the deep calibration: it is recovered low (mean ranging from 2.8×10^{-9} up to 3.8×10^{-9}) throughout the range assayed and always overlaps with the Elephantoidae’s RoS estimate. The variation is more sensible over the precision of that estimate, which regularly improves with increasing sequence length: the HPDs span from ~150% of the average for the shortest sequences down to only 25% over 14 kb (90% and full mt; Fig. 5). However, varying the fragment length has a dramatic effect on the posterior RoC when the shallow calibration is used: the average RoC shows a 40-fold drop (from $\sim 4 \times 10^{-7}$ for 1% down to $\sim 1 \times 10^{-8}$ for the full mt) associated with a comparable narrowing of the credibility interval (although the 95% HPDs remain relatively wide throughout the range by reference to the mean rate, i.e., spanning $\geq 200\%$ of

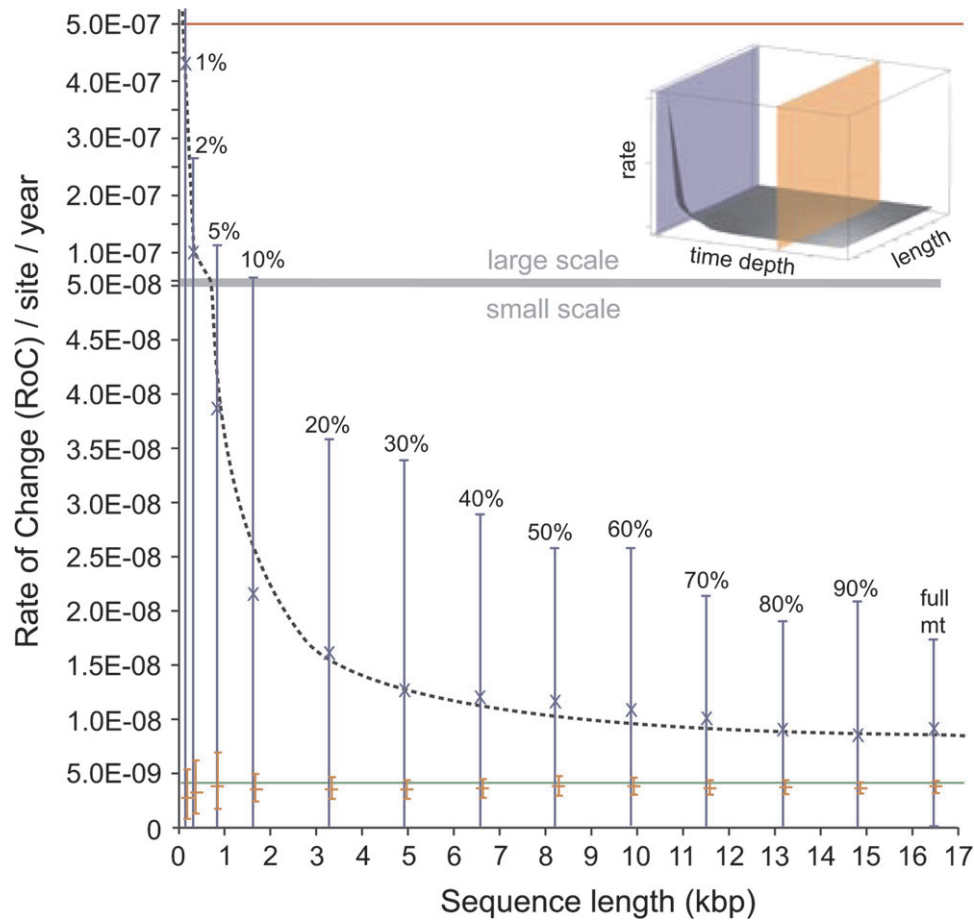


FIGURE 5. Analysis of the posterior RoC, according to the sequence length analyzed. Results (i.e., posterior average and 95% HPD) are displayed for the shallow calibration (in blue) or the deep calibration (in orange, see text) and referred to in the top-right pane of a signal-dependent model. The scale of the figure was split at 5.0×10^{-8} (gray scale demarcation) to improve the display of both analytical series together. The hyperbolic fit of the full Bayesian analysis is displayed by the dashed line. The green constant is set at the previously published phylogenetic rate for the full mt genome of Elephantioidea: 4.2×10^{-9} substitution per site per year (Rohland et al. 2007), whereas the red constant corresponds to the prior rate selected by Ho, Kolokotronis, et al. (2007) in their simulations.

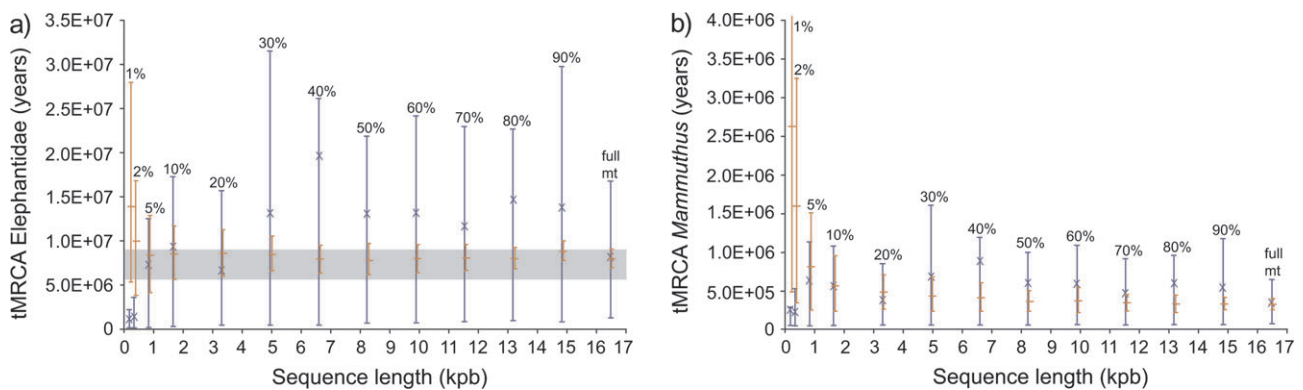


FIGURE 6. Posterior estimates of the tMRCA for the nodes (a) Elephantinae and (b) *Mammuthus*. Average and 95% HPD are displayed consistently with Figure 5. The paleontological estimate for the node Elephantinae is shaded; it is bounded on one side by the differentiation of the Elephantidae (Elephantinae + *Stegotetrabelodon*) approximately 7.5–9 Ma, and on the other side by the earliest evidence of *Loxodonta* and *Elephas* fossils, 5.4–7.3 and 5.2–6.7 Ma, respectively (Tassy 2003; Shoshani and Tassy 2005).

the average estimate). Most of the decrease of the RoC is concentrated in the shortest stages, and when 30% or more (≥ 5 kb) of the genome are analyzed, the average RoC approaches the estimate obtained for the full mt.

This pattern resembles the one recovered in Ho, Phillips, Cooper, et al. (2005)'s analyses of the variation of RoC with time except that the variable incriminated here is the sequence length. Although this result is not anticipated by the time-dependency model (Fig. 3a), it is expected by our hypothesis where the rate acceleration is related to the total divergence between sequences (Figs. 3b and 5). A strict time-dependency hypothesis is therefore likely to be flawed for it only describes 1 dimension of the observations, whereas the hypothesis of a causal role of the AoC can account for the variation of RoC in both dimensions.

The Rate Acceleration Seen as a Sampling Artifact

The appeal of the time-dependency hypothesis lies in its immediate biological significance, whereas the alternative signal-dependency hypothesis will be a disappointment to the biologist for it is not the result of any evolutionary process. It is reasonable to postulate that the results obtained in the analyses of the full mt, whichever calibration is used (shallow or deep), should reflect optimal or quasi-optimal outputs for the genomic data under those priors. Thus, any deviation from these results, that is, when smaller data sets are examined under the same conditions, should be regarded as potential errors. Therefore, we consider the increase (up to 40-fold) observed for the mean RoC under shallow calibration as a computational artifact: the apparent time dependency of the rate acceleration appears as a by-product of a broader signal-dependent bias. The symmetry of our analytical framework (in both time and fragment length; Fig. 3b) also suggests that the low RoCs recovered with deep calibration, even for short sequences, are likely to be the results of an increased phylogenetic content inherent to the calibrations.

If the Bayesian analyses were not biased, random sets of different fragment lengths from the same data matrix should a priori yield comparable mean RoC estimates with a parallel widening of the HPD when the fragment length decreases. This is indeed what we observe for the RoC when we focus on the deep calibration analyses (Fig. 5). With the shallow calibration, however, the widening of the HPD is accompanied by a regular increase in the average RoC when fragment length decreases, so that the error affects not only the precision of the estimate but also its accuracy. This artifact has been previously acknowledged as a potential risk by Ho, Shapiro, et al. (2007, p. 518): "In general, estimates of scale parameters (parameters bounded at zero but with no upper bound) are upwardly biased, whether they are estimated by maximum likelihood or BayesianMCMC. [...] Provided that the sequences are informative and that the dated tips are sufficiently distinct in age, the impact of this estimation bias will be small, but this requires further investigation." Here,

we provide the first experimental evidence of the quantitative impact of this bias caused by the combined effect of short sequences and shallow calibration, which suggests that it has been largely underestimated in the development of the time-dependency hypothesis.

Our results also demonstrate that the deep calibration approach has a large effect against the artifactual variation of the RoC. If the tip calibrations covering only the last 100 kyr are meaningful when the total variation embraced by the alignments is elevated, their calibration power over data sets of very limited information content becomes spurious. By using a deep calibration point, one can compensate for the loss in usefulness of the tip calibrations when short sequences (i.e., low between-sequence divergence) are compared. We name this effect "normalization," as the posterior RoC for deep calibration is centred around the published RoS estimate throughout the sequence range analyzed. This effect carries its own limitation: the RoC derived from any deep calibration can be biased if that deep calibration itself is not accurate. The enforcement of a deep calibration might not be the only way to avoid a signal-dependent artifact. In the approach developed by Emerson (2007), where the width of the range for most priors was tightly enforced in BEAST analyses, similar results were obtained by limiting the sampling of unlikely elevated RoCs in the posteriors.

Modeling the Rate Variation

Even if the rate acceleration observed under shallow calibration is an artifact, modeling its pattern might prove useful to derive a reliable estimate of the RoC. Several regression approaches adequately model the variation of the RoC along with fragment length. Although a good fit is obtained for the vertically translated exponential decay, as measured per least square correlation index ($R > 0.991$), a smoother decay curve based on a (simpler) hyperbolic fit was found to better describe the present data ($R > 0.997$). This model assumes that the decay of the RoC is inversely proportional to the fragment length and converges to a constant minimal rate that can be seen as the RoS:

$$\text{RoC}_{(\text{FgtL})} = \text{RoS} + \frac{\text{Cte}}{\text{FgtL}}$$

The second constant term of the regression (which would represent the instantaneous RoM in the view of Ho, Phillips, Cooper, et al. 2005) has no biological meaning in our hypothesis: it only measures the strength of the artifact in the posterior estimate of the RoC (here $\text{Cte} = 3 \times 10^{-5}$). This equation suggests that if the sequence length was not limited by the size of the mt genome, one could in theory derive a RoC as low as 6.7×10^{-9} rather than the estimate of 9.1×10^{-9} recovered for the full mt data set (95% HPD: 2.3×10^{-10} to 1.7×10^{-8} ; Fig. 5). Despite the long sequence considered (16469 bp), this latter rate can thus be regarded as a slight overestimate of the optimal RoC for the shallow calibration approach. The comparison of this optimal

RoC (6.7×10^{-9}) with the most reliable (i.e., full mt) estimate of the RoC under deep calibration (3.8×10^{-9} ; 95% HPD: 3.3×10^{-9} to 4.3×10^{-9}) suggests that the shallow calibration approach is bound to yield a slightly elevated RoC (in this case less than a 2-fold difference), even after the removal of the signal-dependent artifact.

This residual difference cannot be accounted for by our current hypothesis but might be the results of 2 nonexclusive processes. First, this difference might be explained by different types of errors as proposed by Emerson (2007): 1) the error on the calibration of the terminals, that is, the poorly constrained date for the mastodon sequence (range 50–130 ka, here estimated by its mean 90 ka), will have a stronger detrimental effect on the shallow calibration, 2) any error associated with the calibration of the root at 24–28 Ma could have an impact on the deep calibration approach, 3) sequencing errors or damage, if unable to explain a 40-fold acceleration phenomenon (Ho, Phillips, Cooper, et al. 2005) might be sufficient in some aDNA sequences to explain the residual difference, 4) the mutational saturation between the mastodon and the elephantids might lead to an underestimate of the RoC based on the deep calibration if not properly corrected by the substitution model. Finally, if one excludes all putative sources of error, the difference in the shallow and deep RoCs might reflect a nonerroneous acceleration of the rate observed only when the recent calibrations are used: by essence, a real (although quantitatively limited) time-dependency effect, putatively related to purifying selection (Penny 2005).

CONSIDERATIONS OVER THE PRECISION AND ACCURACY OF tMRCAs

In many aDNA studies, a primary focus is placed on proper dating of significant divergence events rather than estimating the evolutionary rate itself. Little attention has been placed, until now, on the variation of the divergence date estimates that rely on variable RoCs. Figure 6 summarizes the results of the Bayesian analyses for the posterior distribution of the tMRCAs for both the Elephantinae node and the *Mammuthus* node. As the results for the 2 nodes investigated are similar, we will focus essentially on the Elephantinae node (Fig. 6a), which is associated with well-documented paleontological records (Tassy 2003; Shoshani and Tassy 2005) allowing for verification of the posterior estimates.

The first striking result is the consistency between estimates derived from both the shallow and the deep calibration methods when the full mt data set is used. Although the 95% HPDs are smaller when the root is calibrated (as expected a priori), both approaches yield comparable accuracies for the tMRCAs, with close average estimates differing from each other by less than 5%. This result demonstrates that an acceleration of the average RoC (as documented by the ~ 3 -fold difference between shallow and deep calibrations recovered for the full mt data set; Fig. 5) may not cause any symmetrical difference in the tMRCA estimates.

The shallow calibration approach recovers HPDs which overlap with the paleontological estimates when at least 5% (823 bp) of the original sequence length was analyzed (Fig. 6a). However, further increase in the sequence length does not improve the precision of that estimate and generally leads to an overestimated average tMRCA. Conversely, short sequences (1% or 2% of the total mitochondrial DNA) produce more precise but inaccurate (i.e., too recent) dates, which rely on large samples of erroneous topologies in the posteriors where the *Mammuthus* node is not consistently monophyletic, but often a paraphyletic group, which includes modern elephants. This result is counterintuitive in the light of the estimates of the RoC for the same sequence lengths: we show that largely imprecise (and inaccurate) estimates of the RoC can lead to very precise (although inaccurate) tMRCAs.

The deep calibration method generally yields more precise and accurate estimates of the tMRCAs when at least 5% of the original sequence length is analyzed (Fig. 6a). Nevertheless, when sequence length drops to less than 5%, the precision on the tMRCA estimates decreases drastically and the accuracy is reduced (i.e., average tMRCAs are overestimated). Therefore, it appears that enforcing the root calibration prior might transfer a part of the computational artifact (acknowledged for the RoC in the shallow calibration approach) to the posterior estimates of tMRCAs. This issue deserves further investigation beyond the scope of the present paper, but it is worth noting that setting a deep calibration point with a narrow confidence interval in the priors of BEAST analyses does not always lead to recover narrow estimates for other divergences on the tree. Our analyses thus reveal that neither of the 2 calibration approaches examined performs satisfactorily (according to accuracy and precision) for dating divergence events when largely uninformative sequences (i.e., here short sequences) are analyzed.

GENERALIZATION OF THE SIGNAL-DEPENDENT ARTIFACT

Recent publications have begun documenting the differences in RoCs/tMRCAs derived from short versus long sequences under full Bayesian frameworks using BEAST. Here, we use 3 aDNA case studies (mammoth, hominids, and bisons) to illustrate the general significance of the signal-dependent hypothesis.

Case Study 1: Woolly Mammoth Sequence Data

Our first comparison addresses 3 published data sets of woolly mammoth mitochondrial DNA from which discordant RoCs/tMRCAs can now be better appreciated. In the first published paper, Barnes et al. (2007) identified 2 highly divergent clades (referred to as 1 and 2). To estimate their tMRCA, they analyzed 33 sequences (741 bp each) with BEAST according to a full Bayesian framework using only terminal ^{14}C tip dates as calibrations. Under the best-fitting model priors, they recov-

ered an elevated RoC (2.5×10^{-7}) with a wide HPD (6.3×10^{-8} to 4.5×10^{-7}) comparable with the results of Ho, Kolokotronis, et al. (2007) from the same data set (Fig. 1). Based on this RoC, they provide a very young estimate for the tMRCA of Clade 1 at 63 ka (95% HPD: 49–94 ka). This result conflicts with both shallow and deep estimates derived from our Fig. 6b, which support an average coalescence time as old as 320–340 ka for the same node (because all the mammoths used in the present study belong to Clade 1).

Since the original submission of this manuscript, Gilbert et al. (2008) have added 5 complete mt genomes from mammoths documenting both Clades 1 and 2. They also perform BEAST analyses of all available complete genome sequences under a full Bayesian approach using either only ^{14}C date tip calibrations or by enforcing the calibration of the root of their tree using the same divergence estimate as in the current study for the mastodon versus elephantid split (24–28 Ma, although using a lognormal rather than uniform prior). Unfortunately, Gilbert et al. (2008) report neither the actual RoCs from those analyses nor the tMRCA of the Clade 1, which prevents direct comparisons with the previous results. From Figure 3 of their publication (Gilbert et al. 2008, p. 8329), it appears that 1) like in our analyses of the full mt, both shallow and deep calibration approaches yield fairly close tMRCAs estimates and that 2) the average tMRCA of Clade 1 appears much older (~3-fold) than the original estimate by Barnes et al. (2007). Taken together those results suggest that the analyses of Barnes et al. (2007) based on recent calibrations of a 741-bp long alignment have suffered from a signal-dependent artifact, which could be overcome by Gilbert et al. (2008) when using longer sequences.

The most recent paper (Debruyne et al. 2008) uses a larger sample of 138 dated mammoths and capture a more complex phylogeographic picture in which the split between Clades 1 and 2 is still represented (although renamed clades D + E and A, respectively). Not surprisingly (because the fragment length analyzed and the time range of samples were identical), Debruyne et al. (2008) recover RoC/divergence estimates comparable to those of Barnes et al. (2007) when a shallow calibration (i.e., ^{14}C tips only) analysis is performed with BEAST. However, when the posterior estimates of the tMRCAs for the Elephantinae and the *Mammuthus* nodes derived from the full mt analysis (Fig. 6) are implemented as priors to the BEAST analysis, the acceleration is not observed anymore: the posterior RoC (3×10^{-9} ; 95% HPD: 4.9×10^{-10} to 6.9×10^{-9}) is close to the Elephantoida RoS and is some 100 times lower than the estimate by Barnes et al. (2007), resulting in a tMRCA of Clade 1 approximately 5-fold older (average 301 ka; 95% HPD 214–398 ka). Conversely, the estimate of the age of the tMRCA of all mammoth sequences (i.e., coalescence of Clades 1 and 2 at an average 0.89 Ma; 95% HPD: 0.43–1.598 Ma) is fully consistent with the estimate of the same node proposed by Gilbert et al. (2008) using a shallow calibration of the complete mt genomes (1.07 Ma; 95% HPD 0.38–2.43 Ma)—although

slightly more recent than their deep calibration estimate (1.7 Ma, 95% HPD 1.44–1.98). This result confirms that, in spite of a limited sequence length (741 bp), the rate and tMRCAs estimates could be normalized by the use of additional calibration points. This study thus provides a clear framework for most aDNA populational studies that are likely to be limited in sequence length and radiocarbon dated specimens.

Case Study 2: Hominid Sequence Data

Ho, Phillips, Cooper, et al. (2005) have originally analyzed different HVR data sets of hominids that depicted a typical rate acceleration pattern: when calibrated with only ^{14}C dates of Neandertal sequences (data set of 8 terminals, 34 variable positions), the analysis yielded a high average RoC of $\sim 4 \times 10^{-7}$, approximately 10 times as high as the estimates derived from a larger data set (38 terminals, 114 variable positions) calibrated with the human–chimpanzee split set at 4.5–6 Ma.

More Recently, Endicott and Ho (2008) have published an analysis of a large number of complete mt genomes of hominids (177 humans and 2 chimpanzees). They have created different partitions of the data set for which model parameters have been optimized individually. In addition, they have applied 2 different calibration methods: the first one using only radiocarbon-based estimates of haplogroups divergence within humans (<50 ka) and another where the human–chimpanzee split was again calibrated around 5 Ma. Their analyses have revealed an extensive variation of the RoC among partitions (the control region sequences seemingly evolving 3 times faster than the genome seen as a whole). Nevertheless, whatever sequence partition was analyzed, the average RoC showed almost no variation over time: a low ~1.5-fold increase was generally observed between the old and the recent calibrations (Endicott and Ho 2008). Furthermore, for both calibration approaches, the estimates were precise with 95% HPDs spanning ~50% of the average RoC and showing no significant variation when the calibration changed. There is little doubt that the average RoC derived from the analysis of a large number of hominid complete genomes (Endicott and Ho 2008) outperforms the previous estimates based on limited data, again supporting a signal-dependent artifact in the original analysis (Ho, Phillips, Cooper, et al. 2005).

Case Study 3: Ancient Bison Sequence Data

Emerson suggested the authors implement the approach used for the proboscidean data set with the well-studied bison data set (Shapiro et al. 2004). The 2 most distinct sequence sets segregated by calibration range in the study of Ho, Shapiro, et al. (2007) were analyzed separately: 66 sequences spanning 0–10 ka versus 182 sequences spanning 0–60 ka. To address the effect of sequence length, we generated 7 alignments (for each sequence set). Four shorter alignments consisted of a single subset of the original data covering either

one-quarter (154 bp), one-half (308 bp), three-quarters (461 bp), or the entire sequence available (615 bp). In addition, 3 longer alignments were generated through the concatenation of a 2-fold (1230 bp), 5-fold (3075 bp), and 10-fold (6150 bp) of the original data. These last 3 alignments are purely artificial and might not compare with real sequences. The BEAST analyses were performed according to the conditions used by Ho, Shapiro, et al. (2007) on the same data: 30 000 000 generations with a 10% burnin and sampling every 3000 generations, strict molecular clock enforced, HKY + I + G substitution model, and 12 category skyline plot demographic prior.

Figure 7 shows the results of these analyses. If one focuses first on the largest data set (0–60 ka range), a similar pattern as seen for the proboscidean data set is recovered for the bison. When less than the original sequence length is used, the average RoC increases (up to ~3-fold) and its associated error widens from ~50 up to 100% of the mean estimate. For the concatenated alignments, the posterior RoC was not only more precisely estimated than for the 615-bp alignment, but its average estimate was also consistently lower.

The comparison of these results with the ones derived from the short calibration range (0–10 ka, Fig. 7) shows that the signal-dependent artifact affects the latter data set even more heavily: larger increase in RoC average and HPD when the sequence shortens. Like for the real proboscidean sequences, the rate variation for the bison data sets can also be modeled by a hyperbolic fit (Fig. 7). Both hyperbolic fits vary in almost parallel fashion over 5-fold of the original data set and thus never converge to similar estimates: the RoS for the 0–60 ka data set could be as low as 1.2×10^{-7} (rather than the estimate of 3.0×10^{-7} derived from the 615-bp alignment), whereas the RoS for the 0–10 ka range is more than double (2.7×10^{-7}). This difference suggests that fragment length and time depth are not the only

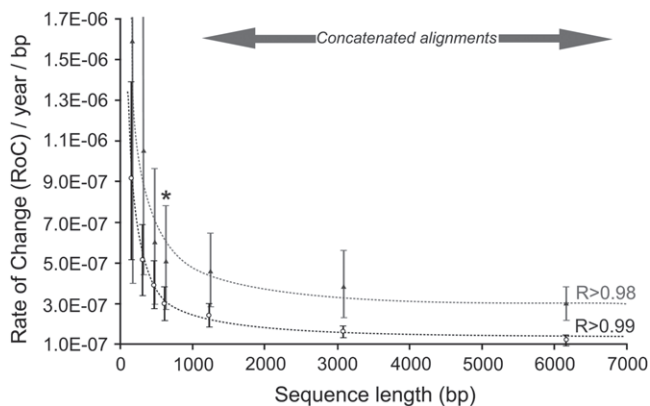


FIGURE 7. Posterior estimates of the RoC (average and 95% HPD) for the re-analyzed bison data spanning (from left to right) 1/4, 1/2, 3/4, 1×, 2×, 5×, and 10× the original data. Both the complete 0–60 ka calibration range (empty circles) and the 0–10 ka range (filled triangles) are displayed. The asterisk indicates the results from the original data without length modification. The hyperbolic fits adjusted to the data are displayed by the dashed lines with associated correlation score.

contributing factors to the rate acceleration. Contrary to the proboscidean study (where the composition in terminals was constant), the 2 bison sequence sets analyzed here differ in the number of terminals and the overall phylogenetic pattern, 2 factors that may contribute to the signal-dependent bias. Alternatively, the difference in RoS estimates could also support a prediction by Emerson that the signal-dependent bias may not be sufficient to explain the entire phenomenon of rate acceleration. In all cases, the approach of the decomposition/concatenation of the original data provides a useful (and relatively easy to implement) tool to evaluate to what extent the RoC derived from any data set limited in both calibration depth and sequence length might be biased due to a signal-dependent artifact.

CONCLUDING REMARKS

Our Bayesian analyses presented here have implications of general significance for aDNA data sets analyzed with BEAST. Based on the proboscidean mt data set, we were able to show that the apparent time dependency of the RoC recovered for inferences built on poorly informative data sets calibrated in time with only recent radiocarbon dates is more likely explained by an artifact than an actual evolutionary paradigm. The limited phylogenetic content of short sequences appears to relax the constraint over the substitution rates, which can vary so greatly that their mean estimate becomes irrelevant for use and leads to a reproducible bias of the apparent acceleration of the molecular rates. By showing how the pattern described from the analyses of the proboscidean data can be extended to other published material, we suggest that all aDNA data sets be tested for such a signal dependency through the analytical framework provided here to evaluate the risk of such a bias.

Once the effects of the signal-dependent artifact are accounted for, the difference between the RoC estimates based on either shallow or deep calibrations suggest that this artifact alone does not account for the entire rate acceleration. It does, however, show that the acceleration phenomenon is certainly of much lower magnitude than has been previously reported by Ho, Phillips, Cooper, et al. (2005).

We have also attempted to address the reciprocal qualities of recent versus deep calibration approaches. The shallow calibration approach is more pertinent to a full Bayesian framework as it relies on the inherent structure and quality of the data to converge to both accurate and precise estimates. However, we have exemplified how this approach can be misleading when the data are poorly structured (Rannala 2002): inaccurate tMRCAs and both inaccurate and imprecise RoCs are then recovered. Conversely, the consistent and accurate results obtained for the RoC with an enforced deep calibration provide support for this methodological approach provided accurate information is available for the enforced calibration(s). Despite the obvious

efficiency of BEAST algorithms in a full Bayesian framework when analyzing long DNA sequences, our analyses suggest that data sets of very limited phylogenetic content might remain out of the range of precise and accurate estimates of divergence dates. The legitimacy of dating divergence events using short, potentially uninformative, ¹⁴C-dated data sets exclusively, is thus of limited value, despite being the norm for aDNA studies.

FUNDING

This work was funded by The Natural Science and Engineering Research Council (grant 299103-2004), the Canadian Research Chairs program, a fellowship from the Human Frontier Science Program (HFSP-00285C/2005), and McMaster University.

ACKNOWLEDGMENTS

We thank Melanie Kuch and Carsten Schwarz for discussions and critical evaluation of our protocol. We are grateful to J. Sullivan, K. Zamudio, B. Emerson, and an anonymous reviewer for their constructive comments. We are grateful to Beth Shapiro and Simon Ho for sharing preformatted files of the bison data set.

REFERENCES

- Bandelt H.J. 2007. Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. *Heredity*. 100:1–2.
- Barnes I., Shapiro B., Lister A., Kuznetsova T., Sher A., Guthrie D., Thomas M.G. 2007. Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr. Biol.* 17:1072–1075.
- Burridge C.P., Craw D., Fletcher D., Waters J.M. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. *Mol. Biol. Evol.* 25:624–633.
- Debruyne R., Chu G., King C.E., Bos K., Kuch M., Schwarz C., Szpak P., Grocke D.R., Matheus P., Zazula G., Guthrie D., Froese D., Buigues B., de Marliave C., Flemming C., Poinar D., Fisher D., Southon J., Tikhonov A.N., Macphee R.D., Poinar H.N. 2008. Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr. Biol.* 18:1320–1326.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Emerson B.C. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst. Biol.* 56:337–345.
- Endicott P., Ho S.Y. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am. J. Hum. Genet.* 82:895–902.
- Felsenstein J. 2005. PHYLIP: Phylogeny inference package. Version 3.6. Seattle (WA): University of Washington.
- Gilbert M.T., Drautz D.I., Lesk A.M., Ho S.Y., Qi J., Ratan A., Hsu C.H., Sher A., Dalen L., Gotherstrom A., Tomsho L.P., Rendulic S., Packard M., Campos P.F., Kuznetsova T.V., Shidlovskiy F., Tikhonov A., Willerslev E., Iacumin P., Buigues B., Ericson P.G., Germonpre M., Kosintsev P., Nikolaev V., Nowak-Kemp M., Knight J.R., Irzyk G.P., Perbost C.S., Fredrikson K.M., Harkins T.T., Sheridan S., Miller W., Schuster S.C. 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl. Acad. Sci. USA.* 105:8327–8332.
- Gilbert M.T., Tomsho L.P., Rendulic S., Packard M., Drautz D.I., Sher A., Tikhonov A., Dalen L., Kuznetsova T., Kosintsev P., Campos P.F., Higham T., Collins M.J., Wilson A.S., Shidlovskiy F., Buigues B., Ericson P.G., Germonpre M., Gotherstrom A., Iacumin P., Nikolaev V., Nowak-Kemp M., Willerslev E., Knight J.R., Irzyk G.P., Perbost C.S., Fredrikson K.M., Harkins T.T., Sheridan S., Miller W., Schuster S.C. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science.* 317:1927–1930.
- Hall T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* 41:95–98.
- Hauf J., Waddell P., Chalwatzis N., Joger U., Zimmermann F.K. 2000. The complete mitochondrial genome sequence of the African elephant (*Loxodonta africana*) phylogenetic relationships of Proboscidea to other mammals and D-loop heteroplasmy. *Zoology.* 102:184–195.
- Ho S.Y., Kolokotronis S.O., Allaby R.G. 2007. Elevated substitution rates estimated from ancient DNA sequences. *Biol. Lett.* 3:702–705.
- Ho S.Y., Phillips M.J., Cooper A., Drummond A.J. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22:1561–1568.
- Ho S.Y., Phillips M.J., Drummond A.J., Cooper A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Mol. Biol. Evol.* 22:1355–1363.
- Ho S.Y., Shapiro B., Phillips M.J., Cooper A., Drummond A.J. 2007. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* 56:515–522.
- Howell N., Howell C., Elson J.L. 2008. Molecular clock debate: time dependency of molecular rate estimates for mtDNA: this is not the time for wishful thinking. *Heredity.* 101:107–108.
- Howell N., Smejkal C.B., Mackey D.A., Chinnery P.F., Turnbull D.M., Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.* 72:659–670.
- Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.
- Krause J., Dear P.H., Pollack J.L., Slatkin M., Spriggs H., Barnes I., Lister A.M., Ebersberger I., Paabo S., Hofreiter M. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature.* 439:724–727.
- Lambert D.M., Ritchie P.A., Millar C.D., Holland B., Drummond A.J., Baroni C. 2002. Rates of evolution in ancient DNA from Adelie penguins. *Science.* 295:2270–2273.
- Parsons T.J., Munie D.S., Sullivan K., Woodyatt N., Alliston-Greiner R., Wilson M.R., Berry D.L., Holland K.A., Weedn V.W., Gill P., Holland M.M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat. Genet.* 15:363–368.
- Penny D. 2005. Relativity for molecular clocks. *Nature.* 426:183–184.
- Poinar H.N., Schwarz C., Qi J., Shapiro B., Macphee R.D., Buigues B., Tikhonov A., Huson D.H., Tomsho L.P., Auch A., Rampp M., Miller W., Schuster S.C. 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science.* 311:392–394.
- Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics.* 14:817–818.
- Rambaut A., Drummond A.J. 2007. Tracer. Version 1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>.
- Rannala B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51:754–760.
- Rogaev E.I., Moliaka Y.K., Malyarchuk B.A., Kondrashov F.A., Derenko M.V., Chumakov I., Grigorenko A.P. 2006. Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biol.* 4:e73.
- Rohland N., Malaspinas A.S., Pollack J.L., Slatkin M., Matheus P., Hofreiter M. 2007. Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* 5:e207.
- Shapiro B., Drummond A.J., Rambaut A., Wilson M.C., Matheus P.E., Sher A.V., Pybus O.G., Gilbert M.T., Barnes I., Binladen J., Willerslev E., Hansen A.J., Baryshnikov G.F., Burns J.A., Davydov S., Driver J.C., Froese D.G., Harington C.R., Keddie G., Kosintsev P., Kunz M.L., Martin L.D., Stephenson R.O., Storer J., Tedford R., Zimov S., Cooper A. 2004. Rise and fall of the Beringian steppe bison. *Science.* 306:1561–1565.

- Shoshani J., Tassy P. 2005. Advances in proboscidean taxonomy & classification, anatomy & physiology, and ecology & behavior. *Quat. Int.* 126–128:5–20.
- Shoshani J., Walter R.C., Abraha M., Berhe S., Tassy P., Sanders W.J., Marchant G.H., Libsekal Y., Ghirmai T., Zinner D. 2006. A proboscidean from the late Oligocene of Eritrea, a "missing link" between early Elephantiformes and Elephantimorpha, and biogeographic implications. *Proc. Natl. Acad. Sci. USA.* 103:17296–17301.
- Tassy P. 2003. Elephantoida from Lothagam. In: Leakey M.G., Harris J.M., editors. *Lothagam: the dawn of humanity in Eastern Africa*. New York: Columbia University Press. p. 331–358.

Received 26 September 2008; revised 5 December 2008; accepted 24 March 2009

Associate Editor: Kelly Zamudio